

GenCore version 4.5
 Copyright (c) 1993 - 2000 Compugen Ltd.
 OM protein - protein search, using sw model
 Run on: July 24, 2002, 04:18:22 ; Search time 67.11 Seconds
 (without alignments)
 4542.051 Million cell updates/sec
 Title: US-10-044-807-2
 Perfect score: 9588
 Sequence: 1 MECCRRATPGTLLFLAFL...LKLCQLSQFKSRCCGTGKA 1762

Scoring table: BLOSUM62
 Gapop 10.0 , Gapext 0.5
 Searched: 562222 seqs, 172994929 residues
 Total number of hits satisfying chosen parameters: 562222
 Minimum DB seq length: 0
 Maximum DB seq length: 2000000000
 Post-processing: Minimum Match 0%
 Maximum Match 100%
 Listing first 45 summaries

Database :

SPTREMBL19:
 1: sp_archaea:
 2: sp_bacteria:
 3: sp_fungi:
 4: sp_human:
 5: sp_invertebrate:
 6: sp_mammal:
 7: sp_mhc:
 8: sp_organelle:
 9: sp_phase:
 10: sp_plant:
 11: sp_rodent:
 12: sp_virus:
 13: sp_vertebrate:
 14: sp_unclassified:
 15: sp_virus:
 16: sp_bacteriap:
 17: sp_archaeap:

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2907	30.3	525	4 Q96RW4	Q96RW4 homo sapien
2	2625.5	27.4	766	4 P82987	P82987 homo sapien
3	2283.5	23.8	417	4 Q9BXY3	Q9BXY3 homo sapien
4	1629	17.0	1023	4 Q9ULI7	Q9ULI7 homo sapien
5	1279.5	13.3	1014	5 Q95R33	Q95R33 drosophila
6	1145.5	11.9	2165	5 Q19791	Q19791 caenorhabdi
7	1088.5	11.4	1059	5 P90884	P90884 caenorhabdi
8	1019.5	10.6	1280	11 Q9EPX2	Q9EPX2 mus musculu
9	994	10.4	192	11 Q9CX59	Q9CX59 mus musculu
10	978	10.2	1235	4 Q95428	Q95428 homo sapien
11	887	9.3	2167	5 Q76840	Q76840 caenorhabdi
12	882	9.2	2174	5 Q9GGR0	Q9GGR0 drosophila
13	882	9.2	3060	5 Q9VAV4	Q9VAV4 drosophila
14	865	9.0	1572	5 O44938	O44938 haemochus
15	857.5	8.9	3198	5 Q9U8G8	Q9U8G8 manduca sex
16	848.5	8.8	761	6 Q951Q2	Q951Q2 macaca fasc

ALIGNMENTS

RESULT 1

Q96RW4	1427	8-8	4	Q96L37
18	761	7-9	5	Q9V9Q7
19	732.5	7-6	4	O60345
20	594	6-2	4	Q9H8X0
21	584.5	6-1	4	Q9NSJ8
22	566.5	5-9	4	Q9UFZ4
23	531	5-5	5	Q9W493
24	508.5	5-3	5	Q96RW7
25	506.5	5-3	4	Q9H8E4
26	506.5	5-3	4	Q96SC3
27	495	5-2	4	Q9C0I4
28	472.5	4-9	5	Q9VF61
29	468.5	4-9	4	O17591
30	458	4-8	6	Q9XSV8
31	452	4-7	4	Q9H8H1
32	442	4-6	4	Q9UPZ6
33	413.5	4-3	4	O17591
34	388	4-0	5	Q95V51
35	382.5	4-0	340	11 Q91256
36	376	3-9	4	Q9H3V5
37	373	3-9	788	5 Q22631
38	352.5	3-7	5198	5 Q76518
39	351.5	3-7	1252	4 Q96DN3
40	349	3-6	1121	5 Q966P9
41	349	3-6	1148	5 Q966Q0
42	347	3-6	1062	5 O19204
43	342.5	3-6	432	4 Q9NPM2
44	332	3-5	446	4 O43384
45	328.5	3-4	235	11 Q91YP4

Q96RW4 PRELIMINARY; PRT; 525 AA.
 AC Q96RW4;
 DT 01-DEC-2001 (TREMBLrel. 19, Created)
 DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)
 DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
 DE ADAM-TS RELATED PROTEIN 1.
 GN ADAMTSR1.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_taxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Hirohata S., Apte S.S.;
 RT "A novel member of ADAM-TS related gene, ADAM-TSRI (A Disintegrin-like
 RT And Metalloproteinase domain with Thrombospondin type I modules
 RT Related gene-1).";
 RL Submitted (AUG-1999) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AF176313; AAK84170.1; -;
 SQ SEQUENCE 525 AA; 58351 MW; 3F180C3CAAA7BA68 CRC64;

Query Match 30.3%; Score 2907; DB 4; Length 525;
 Best Local Similarity 99.6%; Pred. No. 3.2e-204;
 Matches 523; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
 QY 1 MECCRRATPGTLLFLAFL...LKLCQLSQFKSRCCGTGKA 1762
 Db 1 MECCRRATPGTLLFLAFL...LKLCQLSQFKSRCCGTGKA 1762
 QY 61 LSSKSCGRNIRYTCNVDCPPAGDFRAQCCSAHNDVKHHGQFYEWLPVNDPDPNCS 120
 Db 61 LSSKSCGRNIRYTCNVDCPPAGDFRAQCCSAHNDVKHHGQFYEWLPVNDPDPNCS 120
 QY 121 LKQAKGTTTIVELAPKVLDTGTCRYTESLDMCISGLCQIVGCDHQLGSTVKEDNCVCNG 180
 Db 121 LKQAKGTTTIVELAPKVLDTGTCRYTESLDMCISGLCQIVGCDHQLGSTVKEDNCVCNG 180

QY	181	DGSTRVLRGQYKQSLSATKSDDTVAIPYGSRRHRLVLKGPDPHLYLETKTLOGTGKNS	240
Db	181	DGSTRVLRGQYKQSLSATKSDDTVAIPYGSRRHRLVLKGPDPHLYLETKTLOGTGKNS	240
QY	241	LSSTGTFLVNSSVDFOKFPDKELRMAGPLTADFIKIRNSGSADSTVQFIYQPIIHR	300
Db	241	LSSTGTFLVNSSVDFOKFPDKELRMAGPLTADFIKIRNSGSADSTVQFIYQPIIHR	300
QY	301	WRETDFFPCSTCGGQYLSAECYDLRSNRVADQYCHYYPENIKPKKLOECNLDPCP	360
Db	301	WRETDFFPCSTCGGQYLSAECYDLRSNRVADQYCHYYPENIKPKKLOECNLDPCP	360
QY	361	ASDYKQIMPYDLYHPLPWEATPTWATSCSSCGGQISRAVSCVEEDIQGHVTSVEEWC	420
Db	361	ASDYKQIMPYDLYHPLPWEATPTWATSCSSCGGQISRAVSCVEEDIQGHVTSVEEWC	420
QY	421	MYTPKMPAQPCNTFDCPKWLAQWSPCTVTCGGGLRYRVVLCIDHRGMHTGGCSPKTKP	480
Db	421	MYTPKMPAQPCNTFDCPKWLAQWSPCTVTCGGGLRYRVVLCIDHRGMHTGGCSPKTKP	480
QY	481	HIKEECIVPTPCYKPKELPVEAKLPWFKAQLEEGAAVSEEPS	525
Db	481	HIKEECIVPTPCYKPKELPVEAKLPWFKAQLEEGAAVSEEPS	525
RESULT	2		
ID	P82987	PRELIMINARY;	PRT; 766 AA.
AC	P82987;		
DT	01-JUN-2001 (TremBLrel. 17, Created)		
DT	01-JUN-2001 (TremBLrel. 17, Last sequence update)		
DE	01-OCT-2001 (TremBLrel. 18, Last annotation update)		
DE	ADAM-TSL3 PRECURSOR (FRAGMENT).		
GN	ADAMTSL3.		
OS	Homo sapiens (Human).		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.		
OX	NCBI_TaxID=9606;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RA	Hirohata S., Anand-Apte B., Seldin M., Apte S.;		
RT	"Punctin, a member of a new family with similarities to ADAM-TS		
RT	proteases, is a component of extracellular matrix of skeletal		
RT	muscle.";		
RL	Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.		
CC	-1- SIMILARITY: CONTAINS 6 TSP TYPE-1 DOMAINS.		
DR	EMBL; AF237652; AAK15041.1; --		
DR	InterPro; IPR000884; TSP1.		
DR	SMART; SM00209; TSP1; 7.		
DR	PROSITE; PS50092; TSP1; 1.		
KW	Repeat; Signal.		
FT	SIGNAL 1 26	POTENTIAL.	
FT	CHAIN 27 766	ADAM-TSL3.	
FT	DOMAIN 79 123	TSP TYPE-1 1.	
FT	DOMAIN 422 474	TSP TYPE-1 2.	
FT	DOMAIN 482 528	TSP TYPE-1 3.	
FT	DOMAIN 568 625	TSP TYPE-1 4.	
FT	DOMAIN 648 703	TSP TYPE-1 5.	
FT	DOMAIN 707 759	TSP TYPE-1 6.	
FT	NON_TER 766 766		
SQ	SEQUENCE 766 AA; 85987 MW; A42613E87AE91719 CRC64;		
Query Match	27.4%;	Score 2625.5; DB 4; Length 766;	
Best Local Similarity	62.5%;	Pred. No. 2.3e-183;	
Matches 448; Conservative	104; Mismatches 144; Indels 21; Gaps 4;		
QY	22	SSRTARSEEDRDLGWDAGPWSCEKTCGGGASYSLLRCLSSKSCGRNIRYTCNSVDC	81
Db	64	TSRNTSRDEKDGNDWAGWDSCSRCTCGGASYSLLRCLTGRNCRGQNIRYKTCNSHDC	123

QY	82	PPERAGDFRAQCSAHNDVKHHGQFYEWLPVSNDDPNPCLSKCOAKGTTLVVLEAPKVLGD	141
Db	124	PPDAEDFRAQCQSAINDVQYQGHYEWLPYNDPAAPCALCKHAQGGVWVVELAPKVLGD	183
QY	142	TRCYTESLDMCISGLQIVGCDHOLGTVKEDNCVCGNDGTCRLVRGQYKQSISAFKS	201
Db	184	TRCNTSLDMCISGLQIVGCDHOLGTVKEDNCVCGNDGTCRLVRGQYKQSISAFKS	243
QY	202	DDTVVAIPYGSRRHRLVLKGPDPHLYLETKTLOGTGKNSLSTGTFLVNSSVDFOKFPD	261
Db	244	EENIATVPLGSRVRIITVKGPVHLFIESKTLOGSKGEHSFNSPGVVFVVENTTVEFORGE	303
QY	262	KEILLRMAGPLTADFIKIRNSGSADSTVQFIYQPIIHRWRETDFPCSACTGGGYOLTS	321
Db	304	QRTKIPGLFADFIKIRYTAAKDSVVQFFYQPIISHOWQTDFFPCTVTCGGGYOLTS	363
QY	322	AECDYDLRSNRVADQYCHYYPENIKPKKLOECNLDPCPASDGYKQIMPYDLYHPLPWE	381
Db	364	AECDYDLRSNRVADQYCHYYPENIKPKKLOECNLDPCPASDGYKQIMPYDLYHPLPWE	423
QY	382	ATPWTACSSCGGQISRAVSCVEEDIQGHVTSVEEWCMTYTPKMPAQPCNIFDCPKWL	441
Db	424	HNPTACSVSCGGIQRKSFVCEESMHGELQVBEWCMYAPKPKVMQTCNLFDCPKWI	483
QY	442	AEQWSPCTVTCGGGLRYRVVLCIDHRGMHTGGCSPKTKPHKEECIVPTPCYKPKELPV	501
Db	484	AMWQSCTVTCGRGLRYRVVLCINRGEHVGGCNPOLKHIKEECVPIPCYKPKELPV	543
QY	502	EAKLPWFKAQLEEGAAVSEEPSFIPEAWSACTVTCGVGTQVIRIVRCVLLSFSQVAD	561
Db	544	EAKLPWFKAQLEEGAAVSEEPSFIPEAWSACTVTCGVGTQVIRIVRCVLLSFSQVAD	603
QY	562	LPIDECEGPKPASQ-----ACVAGPCSGEIEPFNDETDGLGGLQDDELVDYEGF	616
Db	604	LPDECEGPKPLTERPCILLEACDESPASRELDIPLP-----EDSETTYDWEYAGF	653
QY	617	TKCSGCGGQYFAVVSCLNKTREPARENLGVTSRRPPQLKSCNLDPCPARWEIKGS	676
Db	654	TPCTATCLGGHQAIAVLIHIOQTQVNDSLCDMVRPPAMSGACNTEPCPPRHHVGSNG	713
QY	677	PCSLTCGVGLQTRDFVFCSHLLSRENNETVILADELCRQPKPSTVQACNRFNCPAWY	733
Db	714	PCSATCGVGIGTRDYYCLH-----PGETPAPPEE-CRDEKPHALQACNFDCCPGWH	764
RESULT	3		
ID	Q9BXY3	PRELIMINARY;	PRT; 417 AA.
AC	Q9BXY3;		
DT	01-JUN-2001 (TremBLrel. 17, Created)		
DT	01-JUN-2001 (TremBLrel. 17, Last sequence update)		
DT	01-OCT-2001 (TremBLrel. 18, Last annotation update)		
DE	THROMBOSPONDIN.		
OS	Homo sapiens (Human).		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.		
OX	NCBI_TaxID=9606;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RA	Mao Y., Xie Y., Zhou Z., Zhao W., Zhao S., Wang W., Wang Y., Wang S.,		
RA	Tang R., Chen X., Wu C.;		
RL	Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.		
DR	EMBL; AF251058; AAK34948.1; --		
DR	InterPro; IPR000884; TSP1.		
DR	Pfam; PF00090; tsp_1; 3.		
DR	SMART; SM00209; TSP1; 5.		
SQ	SEQUENCE 417 AA; 46274 MW; EF82550656556DC9 CRC64;		
Query Match	23.8%;	Score 2283.5; DB 4; Length 417;	
Best Local Similarity	97.6%;	Pred. No. 9.8e-159;	
Matches 407; Conservative	3; Mismatches 6; Indels 1; Gaps 1;		


```

RL Nature 368:32-38(1994).
DR EMBL; Z69361; CAA93288.1; -.
DR EMBL; Z69360; CAA93288.1; JOINED.
DR EMBL; Z69360; CAA93287.1; -.
DR EMBL; Z69361; CAA93287.1; JOINED.
DR HSSP; P15167; 1DTH.
DR MEROPS; M12.135; -.
DR InterPro; IPR001590; Reprolysin.
DR InterPro; IPR000884; TSP1.
DR InterPro; IPR000130; Zn.Mtpeptidse.
DR Pfam; PF01421; Reprolysin; 1.
DR Pfam; PF00090; Tsp.1; 14.
DR SMART; SM00209; TSP1; 18.
DR PROSITE; PS02015; ADAM_MERO; 1.
DR PROSITE; PS00092; TSP1; 6.
DR PROSITE; PS00142; ZINC_PROTEASE; UNKNOWN.1.
SQ SEQUENCE 2165 AA; 244397 MW; FCC3DA8AAA9C4888 CRC64;

Query Match 11.9%; Score 1145.5; DB 5; Length 2165;
Best Local Similarity 22.1%; Pred. No. 2.6e-74;
Matches 414; Conservative 185; Mismatches 553; Indels 723; Gaps 74;

QY 33 DGLMDAMPWSECSRTCCGGASYSLLRCLS-----SKSCGERNTRYRTCSNVDCPPPEAG 86
Db 603 DQWGDWRSWGECSRTCCGGVQKGLRDCDSPKPRNGGKYCQGRERYRSCNTQCECPWDTQ 662

QY 87 DFRAQOCSAHN--DYKHHG---QFYEWLP--VSNPDNPCLCKQAKGTTLVVELAPKVL 139
Db 663 PYREVQCSSEFNKDIGQVASTNTHWPKYANVAPNERCKLYCLRSGSAFYLLRDVW 722

QY 140 DGTTCYTESLDMCTSLGCOIVGCHQSLGTSVKEDNCGVNGDGGTCRLVRGOYKSLSAT 199
Db 723 DGTCDNRGDDICVAGACMPACQHLSTLRDKCKCGCGDSSCKVVKGTENEQ--CT 780

QY 200 KSDDTVAIPYGSRHIRLVLAG-----PDHLYLETKILQGTKGNSLSSTGTFVLVDSNV 254
Db 781 FGYNEVKIPAGSANIDIRQYGNMKEDNYL---SLRAANGFLLNGHFQVSLARQOI 837

QY 255 DFQ-----KPPDKELLRM--AGPLTADIVKIRNSGASDSTVQFIF-----YQPI 297
Db 838 AFQDTVLEYSGSDAIIERINGTPIRSDIYVHVLVSGSHPPDISVEYMTAAVPAVNIPI 897

QY 298 ---LHRMRETD-FTPCSATCGGGLQLSAECYDLRSNRVADQYCHYYPENI-KPKPKLQ 352
Db 898 SSALYLNRVTDITWTECDRAC-RGOOSOKMLCLDMSTHRQSHDRNC---QNVLPKQATR 952

QY 353 ECNLD-----PCPASDGYKQ-----IMPVDLYHLP----- 378
Db 953 MCNIDCSTRWTTEDVSSCAKSGQKQRQVSCVKMEGDRQTPASEHLCDRNSKPSDIAS 1012

QY 379 -----RWEATPWATACSSCG--GGTOSRAVSCVEEDIQGHVTSVFEWKCMYTPKMPIA 429
Db 1013 CYIDCSGRKKNYGEWTSCTSETCGSNGMKIRKSCYVDOSNR---RVDES LGREQKEATE 1068

QY 430 QPCNIEFGKWLQAGWSPCTVTCOGLRYRVVLICD--HRCMHTGGSP-KTKPHIIEEC 486
Db 1069 RECNRIPCRAWYVGHWSECSRSCDGGVKMRHAQCILDAADRETHSRGPAQTOEHCHNEHA 1128

QY 487 IVPTPCYKPKLEPLVEAKLPWFQAOELEGAAVSEEPSFTIPEAWSACTVTCGVGTQVRI 546
Db 1129 CT-----HWQFG-----VWSDCSAKCGDGVQYRD 1152

QY 547 VRQVLLSFQSQSVADLPIDCEGPKPASQACYPASGTEIPFNPDFTDGLFGLQDFD 606
Db 1153 ANCT-----DRHRSVLPDEHRLKMEKIITKPCHRESC---PKY----- 1187

QY 607 ELYDWEYEGFTKSCSECGGVQVAVSLNKKQTRPAEENLCVTSRRPOLLKSCNLDPC 666
Db 1188 KLGFW-----SQCSVSDGCSWRRRSCVSCNGTE-VDMSLCGTASDRPASHOTCNLGT 1241

QY 667 PARWEIGKWSPCSLTCCGVGLQTRDVFCSHLLSRMNETVILADELCRQPK-PSIVQACNR 725
Db 1188 IIFMLEPAVPKCEKLEFPKPNESQTCELNPCD-----SEFK 1885

```

```

Db 1242 PF-WRNTDWSACSVSCGIGHRETTTC---IYREQS-----VDASFCDTKMPTSQTCHL 1293
QY 726 FNCPPAWYPAQWQPCSRCTCGGVQKREVLCQKQMAQSGFLELPETCSA-SKPACQACK 784
Db 1294 LPC-TSWKPSHWSPCSVTCGSGIQTRSVSC-TRGSEGTIVD--EYFCDNRNTRPLKKTCE 1349
QY 785 KDDC-----PSEWLLSDWTECSTSCGSGTQTRSAICRKMMLKTGLSTVYVNST 830
Db 1350 KUTCQDGPVYLQKLOADVPPIRWATGFWTACSAFCGNGTORLLKCRDHVRD-----LPDE 1404
QY 831 LCPPILPFSSTIRPCMIAATCARFPRSTKSHPHIAAARKVYIQTTRQRKLFHVVGFPAYLL 890
Db 1405 YCNHLDKEVSTNCRLRDCS-----
QY 891 PKTAVVLRCPARRVRKPLITWEKDGQHLSSHTVTV--APFGYLKTHRLKPSDAGVYTCS 948
Db 1426 WKMAEWEECPA-----TCGTHVQOQRNVTVCVSAEDGGRTI--LKDVCDVQ--- 1469
QY 949 AGPAREHFVIKLIIGNRKLVARPLS-----PRSEEEVLAGRKGKGGKEALQTHKHQN 999
Db 1470 -----KAPTSAARCRLEPCPAGEEHI-----GS 1492
QY 1000 GIFSGNSKAEKGLAANPGSRDYDDLVSRLLEQGWPGELLASWEAQDSARNTTSEDPG 1059
Db 1493 WIIGDWSKC-----SASCG-----GGW-----RRRSVSCITSSS 1520
QY 1060 AEQVLLHLPFTVMTQORLDDILGNLSQOPELRDLYSKHLVAQLAEITFRSHLEHQDTL 1119
Db 1521 CDE-----TRKPKMFD----- 1531
QY 1120 LKPSERRTSPVT-----LSPHKHVSFSSSLRTSTGDAAGGSRPRHRKPTILRKISAAQ 1174
Db 1532 -KCNEELCPPLTNNSQWISPTWTHCS-----VSCGGGVQR-----RKICWD 1571
QY 1175 QLSASEVTHLGQTVVALASGTLVLLHCEAIGHPRPTISWARGNEVOFSDRILLQPDSD 1234
Db 1572 VLS-----GRKQDDIECSE--IKPRE- 1590
QY 1235 LQILAPVADYGFYTCNATNALGYSVSIATVLACKPLVKTSRMVIVTEKPAVTVDIGS 1294
Db 1591 -----ORDCEMPPCR----- 1600
QY 1295 TIKTVQGVNVTINCOVAGVPEAEVTFERNKSKLGSPPHILHEGSLLTINVSDDOGLYSCR 1354
Db 1601 -----SHYANKTSSASMTSLSSSSNSMTSSA 1626
QY 1355 AANLHGETESTOILLIDPPQVPTOLEDIRALLAATGPNLPVSLTSPGLTQLVLDPGNSA 1414
Db 1627 S-----ASSLPILP----- 1635
QY 1415 LLGCPKIGHVPVNIWPHGGOPIVTATGLTHHILAAGQILQVANLSCGSGOGEFSCLAONE 1474
Db 1636 -----PWVS----- 1639
QY 1475 AGVLMQKASLVIODYWNSVDRLATCSACGNRGVQOOPRLRCLLNSTEVNPA--HCAGKVR 1532
Db 1640 -----WQTSAWSACSAKCG-RGYKRRVVECVNPSLVNTVASTEDCDQPK 1682
QY 1533 PAVOPIACNRDRCPSRMWTWSACTRSCGGVQVTRRVTCOKLKASGITSPVSNDMCTQV 1592
Db 1683 P-VEEVRCRTKHCP-RWKTITWSSCVTCGRGIRREVOQYR---GRKNLVSDSECNPK 1736
QY 1593 AKR-----PVDYQACNQQLCVWAFSSWQCGNGPCIGPHLAVOHQVFCOTRDGITLP 1645
Db 1737 TKLNSVANCFFV---ACP---AYRWNTVTPWSCKCKDECAQO--KQTRRVHCISTSGKRAA 1788
QY 1646 SEQCSALPRPVSTQNCWSEACSVHVRVSLWTLCTATCGNYGFSQSRVEEC----- 1694
Db 1789 PRCELARAPTISIRECDTSNCPYEWFGDQWQTSKSGCE-GVQTRVRCRRKKNFNSTIP 1847
QY 1695 --VHARTNKAVPEHLCWSGPRPANORCNITPCENMECDRTTRYCEKVQKQLKQLSQFK 1752
Db 1848 IIFMLEPAVPKCEKLEFPKPNESQTCELNPCD-----SEFK 1885

```

```

QY 1753 -----SRCCGTCCK 1761
Db 1886 WSGFGWGECSKNCQG 1900

RESULT 7
P90884
ID P90884 PRELIMINARY; PRT; 1059 AA.
AC P90884;
DT 01-AUG-1998 (TrEMBLrel. 07, Created)
DT 01-JAN-1999 (TrEMBLrel. 09, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE F53B6.2 PROTEIN.
GN F53B6.2.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Pelodierinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RA White S.;
RL Submitted (OCT-1996) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=99069613; PubMed=9851916;
RA none;
RT "Genome sequence of the nematode C.elegans: A platform for
RT investigating biology.";
RL Science 282:2012-2018(1998).
DR EMBL; Z81086; CAB03121.1; .
DR InterPro; IPR003598; Ig_c2.
DR InterPro; IPR003006; Ig_MHC.
DR Pfam; PF00047; Ig; 1.
DR SMART; SM00408; IGC2; 1.
DR SMART; SM00209; TSP1; 2.
DR PROSITE; PS00092; TSP1; 2.
KW Immunoglobulin domain.
KW SEQUENCE 1059 AA; 117768 MW; 5B182A42E41C8597 CRC64;

Query Match 11.48; Score 1088.5; DB 5; Length 1059;
Best Local Similarity 19.38; Pred. No. 1.3e-70;
Matches 350; Conservative 188; Mismatches 446; Indels 827; Gaps 50;

QY 11 TLLFLAFLLSSRTARGEEDRGL-WDAGFPWSCSRCTCGGASYSRLRCLSSKSCGR 69
Db 6 TWVFLFLVILLASHV-----DALSWAASFPWSSCTKTCGGVSRQLRCLTSK-CSGE 58

QY 70 NIRYTCNVDPPEAGDFRAQCSAHNDVKIHGQFYEWLPVSNPDNPCLSKCAQGT 129
Db 59 SVREKVCQAQKTESKSLRDLTICGG-BEIVSRGQ-----CEVVCRLSLTG 103

QY 130 LVVELAPVLDGTRY--TESLDMCTSGICQIVGCDHQLGTSVKEDNCVGCNGDGTCLIV 188
Db 104 --ANFLWRVDDGTPCQAATSRAVCSKSGCQIVGCCDLISSFRFACGVCVGGGQDTC--- 158

QY 189 RGQYKSQLSATKSDDTVAIPYGSRHRLVKGPDHLYETKLTQGTGENSLSTGTLF 248
Db 159 ----- 158

QY 249 VDNSSVDFQKFPDKELLRMAGPLTADFIVKIRNSGSADSTVQFIYQPIIHKWRETDFFP 308
Db 159 -DNGK-----FIWKV-----SEETFA 173

QY 309 CSATC-----GGGYQLTSAB-----CYDLRSNRVADQYCHYPENIKPKQLQECNL 356
Db 174 CASNCDIDVWSGAGRSTASTSQPIVVCVNAITGRVVPKLC---ADKLKPKVEARPCPM 230

QY 357 DPCPSADGKYQIMPYLDYHPLPRWEATPWATCSSCGGGIQSRAYSCVB--EDIQGHVTS 414
Db 231 LICPS-----RWMAADWTECVPHCGEGTRKREVYCVQTAHNVTVH--- 270

QY 415 VEWKCMYTPKMPJAQCNIEFDCPKWLAQWSPCTVTCGQGLRYRVVICIDHRGMHTGG- 473
Db 271 VPDTCENGTRPAABENCVCSTSCGRWEAGKWKCKTASCGQGVRRRHVACV-----GGS 323

QY 474 -CSPKTKPHIKECIVPTPCYKPKKLPVEAKLPWFKAQAELEGAAYS-----EBPS 525
Db 324 DCEGGRPRQETTCYAGIPCSSTATN-----SLDWNDRAY-LDGNFTGSMNDHNDWQAPR 376

QY 526 FIEPAWSACTVTCGVGTQVRIVRCQVLLSFQSQVADLPIDECEG-PKPASQACVAGPC 584
Db 377 LVAGEMTSCSTCGVMSRIVECVAVNPISAPIKLPMSQCQDQPKLFESCEVRSC- 435

QY 585 GEIPEFNDETDLGEGLODFDELVDWEYEGFTKSCSGGGGVQVAVVSLKNQRTREAE 644
Db 436 -----PLQED---SKLSEDEAPYQWRYGDMWTQCSASCLGKGKKAALKCIQVSTGKSQ 485

QY 645 ENLCVTSRRPOLLKSCNLDPCPARWEIGKWPSCSITCGVGLQTRDVFCSHLSRE--MN 702
Db 486 WSOCDARRRPPKSRPCNQHPCCPFWLTSKYSDCSMSCSGGTARRSVKCAQTVSKTDGAD 545

QY 703 ETVILADELCQPKPSTVQACNRFNCPFAWY-----PAQWPCSRCTCGG 746
Db 546 AHIVLRDDRCHFKPKQETETCNVAVACPATWVSSLNKRHNKIKLNKLTQAQWTECSRCD 605

QY 747 GVQKREVLCORMADGSLFLELPTFCSA-SKPACQOQACKDDCPSEWLLSDWTECSTCG 805
Db 606 GERRQVMECEIRDSRGKTQRREDVEDCANTKP----- 637

QY 806 EGTQTRSAICRKMLKLTGLSTVVNSTLCPPLPFSSSIRPCMLATCARPGPSTKHSPIAA 865
Db 638 -----QIVEVCSFGSCSRPE-----LL 654

QY 866 ARKVVYIOTRRQRLHFEVVGGFAYLLPKTAVLRCRPARYRKPLITWEKDGHLISHTVT 925
Db 655 SNRVEQNAEOKKLTJLIGGVAILYQGTISIKKCPAKKFDKKIYKWKNGKKIKNDARIK 714

QY 926 VAPFGYLKHLRLKPSDAGVYTCAGPAREHFKVILGGRNKLVARPLSPRSEEVLAGRK 985
Db 715 VSANGNLRYFHARMEDAGYEC----- 736

QY 986 GGPKEALQTHKQNGIFSNKGAEKRLAANPGSRYDDLVSLLEQGWGPGLLASWEAQ 1045
Db 737 ----- 736

QY 1046 DSAERNITSEEDPGAQVLLHLPLFTMTVEQRRLDDILGNLSQOPEELRDLYSKHLVAQLA 1105
Db 737 -----FT-----DRLOQNVN----- 746

QY 1106 QELFRSHLEHODTLKPSERRTSPVTLSPKHVSGFSSSLRTSSGTGADGGGSRPRPKPT 1165
Db 747 -----LNFKYRDF---PASR----- 758

QY 1166 ILRKISAAQOLSASEVVTWHLGQTVALASGTLVLLHCEAIGHPRPTISARNGEEVQFSD 1225
Db 759 -----VDLA-----PKQIPSTKNRQVQVSK 780

QY 1226 RILLQDDDSLIQILAPVEADVGYTCNATNALGYDSVIAVTLGKPLVKTSMVTINTEK 1285
Db 781 EDVLRQASV----- 790

QY 1286 PAVTVDIGSTIKTVQGVNVTINQVAGVPEAEVFWPRNKSGLSPHHLGSLLLNVSS 1345
Db 791 -----LHK-----MNVS 798

QY 1346 SDQGLYSCRAANHGBELTESTQLLILDPQVPTQLEDIRALLAATGNPLFSLVTSPLGTQ 1405
Db 799 -----ICALL--TAPN----- 807

QY 1406 LVLDPGNSALLGCPKIKHPVPNTWFGHGOPIVATGLTHHILAAAGIIOVANLSGSGSQ 1465
Db 808 ----- 807

```

```
QY 1466 EFSCLAQNEAGVLMOKASLVIOQYVWSDRLATCSASCGNRGVQVQPRLLKLLNSTEVNPA 1525
Db 808 -----DEKARQLRSY-----GNELV----- 823
QY 1526 HCAGKVRPAVQPIACNRDRCPGRWMTWSAC-TRSCG-GGVQTRRVTCOKLKASGISTP 1583
Db 824 -----ARWDLGHWSERQKQICHVAGVQARGISC-KVTFHGEIRN 861
QY 1584 VSNMCTQVAK-RPVDYTOACNOOLCVENAFSSWGQCNCP-CICPHLAVOHROVFCQTRDG 1641
Db 862 VDNSTICESLASVRPPEIRPCHREDCPRWEASQWSECSQRQVSSMLAQKRRNVTCRFNG 921
QY 1642 IITLPEQCSALPRPVSTONCHSEACSVHVRVSLWTLTLCATCGNYGFSQRVCEVHARTNK 1701
Db 922 TSDVLOHCDITNRPATTMDCPNQNCKAMRTSDMGSCSECGTGGVQLRLLSVCYWISSGR 981
QY 1702 AVPEHLCSWGRPRANWQRC-----NITPC-----ENMECRDTRVCEKVQLKLQOL 1748
Db 982 PAGNR-CEQMRPHSARCAVDEPLPPCMPTASALYQRDASCQDSRFRCDIILKLFHSCDS 1040
QY 1749 SQFKSRCCGTC 1759
Db 1041 LEVRQKCCSTC 1051

RESULT 8
Q9EPX2 PRELIMINARY; PRT; 1280 AA.
AC Q9EPX2;
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DE 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE PAPILIN.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
ON NCBI_TaxID=10090;
RX MEDLINE=20530499; PubMed=11076767;
RA Kramerova I.A., Kawaguchi N., Nelson R.E., Fessler L.I., Chen Y.,
RA Kramerov A.A., Kusche-Gullberg M., Kramer J.M., Ackley B.D.,
RA Sieron A.L., Brockop D.J., Fessler J.H.;
RT "Papilin in development; a pericellular protein with a homology to the
RT ADAMTS metalloproteinases.";
RL Development 127:5475-5485(2000).
DR EMBL; AF314171; AAC41980.1; -.
DR HSSP; P12111; 2KNT.
DR InterPro; IPR003599; Ig.
DR InterPro; IPR003598; Ig_c2.
DR InterPro; IPR003600; Ig_like.
DR InterPro; IPR003006; Ig_MHC.
DR InterPro; IPR002223; Kunitz_BPTI.
DR InterPro; IPR000884; TSPI.
DR Pfam; PF00047; Ig; 3.
DR Pfam; PF00014; Kunitz_BPTI; 1.
DR Pfam; PF00090; tsp_1; 5.
DR PRINTS; PR00759; BASICPTASE.
DR SMART; SM00409; IG; 3.
DR SMART; SM00408; IGC2; 3.
DR SMART; SM00410; IG.Like; 2.
DR SMART; SM00131; KU; 1.
DR SMART; SM00209; TSPI; 5.
DR PROSITE; PS00280; BPTI_KUNITZ_1; 1.
DR PROSITE; PS0279; BPTI_KUNITZ_2; 1.
DR PROSITE; PS50092; TSPI; 3.
KW Immunoglobulin domain; Serine protease inhibitor.
SQ SEQUENCE 1280 AA; 138824 MW; AE287705E561AF30 CRC64;
```

Query Match 10.6%; Score 1019.5; DB 11; Length 1280;
Best Local Similarity 24.0%; Pred. No. 1.9e-65;

```
Matches 380; Conservative 153; Mismatches 527; Indels 523; Gaps 62;
QY 16 LAFLLLSRT---ARSEEDRDLWDAPWSECSRTCCGGASYSIRCLSSK-----SC 66
Db 6 LFLSLLLTTPGSWANVRNRQSDTWGTGSEWSPCSRTCCGGISFRERPCYSQRDDGTSC 65
QY 67 EGRNIRYRCSNVDCPEAGDFRAQCSAHNDVKHGFQEWLVPVSNDDPNCSLKCOAK 126
Db 66 VGPARSRTCHTESCPDGYRDFRAEOCAEFDTDFGRRYRWLPYYAAP-NKELNCIPK 124
QY 127 GTTLVVELAPKVDGTRCYTESLDMCISGLCOIVGDHQLGSTVKEDNGVNGDCSTCR 186
Db 125 GONFYKHKDAVDGTPCEPGQDICDGVCRVWGDHKLDSIKQEDKCLQCGDGSSCY 184
QY 187 LVRGYKKSQLSKSDTVAIPYGSRHRLVLKGDPHLYLETKLQ-----TKG 237
Db 185 PVTGFDGN-DISRGYNOIFILPAGATISRIEFAASRNFLAVKISRGYYLNGHWTIEA 243
QY 238 ENSLSSTGTFLVDSVDFQKPFDEILRMAGPLTADEIVKIRNSGSADSTVOIFYQPI 297
Db 244 AQALPVASTVLOYERCVEGDLAPER--IQARPTSEPIVIELL-SQESPGVHYEYILPA 300
QY 298 I-----HRWRETFDFPCSATCGGQYLTSAECYDLRSNRVADAOYCHYYPENIKPKPKIQ 352
Db 301 NDPGRGFSWSHGSGDCSAECGGGHQSRVFC--TIDNEAYPDHMCQHQP---RPTH-R 354
QY 353 ECNLDPCPASDGYKQIMPYDLYHPLPRWEATPWTACSSCGGGIQSRAVCVEEDIQGHV 412
Db 355 SCNTOPCPKT-----KRMKVGIPWTPTCSVCGGGVQSRYSYCISSDTGGO 399
QY 413 TSVEEWKCM-YTPKMPIAQPCNIFDCPKWLAQEWSPCTVTCGGQLRYRVVLCIDHRG--M 469
Db 400 EAAEETOCAGLAGKPTTQACNLQHCANVSEVPWEGCVTCGTGIRKRVTCGRDGSVP 459
QY 470 HTGCGSPYTKPHIKEECIVPTPCYKPKELPVEAKLPWFKAQOELEEGAASVEEPSFIPE 529
Db 460 HAAACLLKQDPTLTPECV-----QEACPVFRGQAWH-----VG 492
QY 530 AWSACTVTCGVGTQVIRVRCQVLLSFSQSVADLPIDEC---EGPKPASORACYAGPCSGE 586
Db 493 SWSLCSKCGGSIIRRRQVCTI-----GPPRCVLDLQSSKPAEAMEACNRQPC--H 540
QY 587 IPEFNPDETDGLGFLQDFDELYDWEYEGFTKCE------SCGGGVQEAIVSC 634
Db 541 LPQEV-----SIQD-----PTRSSDPRLSGPRVSPVSDGRDQQWAPLER 582
QY 635 LNKQT--REPAEENLCVTSRRP-----POLLKS-----CNLDP---CFARW--EIG- 673
Db 583 PRAQSNPREGQDPNLSAGRAPTLQRPHPQPLRPSGPRDCRHSPHGCCPDGHTFSLGP 642
QY 674 KWSPCSLTCGVGLQTRDVFCSHLLS-----REMNETVILLADELC 712
Db 643 QWQCGPLAGASCLASRYCCPDGVSAAEGPOOAGCTRSHGSDNTGNRPGSRVASKNPKI 702
QY 713 RQP-----KPSTVQACNRFNC-----PPAWVP----- 734
Db 703 HQQAHEGEPESECS-SREFGCCYDNVASAAGPLGBCGVQPSYAYPVRCILPSSAGSGCD 761
QY 735 --AQW-----QPCSRTCGG--GVQKREVLCKORMA--D 761
Db 762 WAARWYFVASVGRNRFWYGCHGNANFNASEQECMTCTCGHQHGPFRPEAGAAGRAHVD 821
QY 762 GSFLPELPETFCASAKPACQQAACKDDCPSEWLLSDWTECSTSCGEGTQTSATCKMKLT 821
Db 822 G-----GQRGPGQOE-----PDWHRAGATIPRLPSPGSPWRREQEP 859
QY 822 GLSTVWNSTLCPLPFPSSIRPCM--LATCARPGRSTKHSP-----HTAAARKVYIOTRR 875
Db 860 APCEPPHIPAYGNRPGQEI RPRVPGLDREARPAVPT-HSPSYRIRIAGSEPSLQA-- 916
QY 876 QRKLHFVVGGFAYLLPKTAVVLRCPARRVRKPLITWBDQGHLLISSTHTVIVAPFGYLIK 935
Db 917 -----APQAVQLFCPGNIPSEFQAGWQKEGRP-ISSNRYQLQDGSLLIS 961
```


Matches	371;	Conservative	151;	Mismatches	488;	Indels	526;	Gaps	65;
---------	------	--------------	------	------------	------	--------	------	------	-----

QY	13	LLFLAFLLL-----SSRTARSEDROGLDAMQWPWSECSTRTCGGASYSLRCLSSK--	64
DB	3	LLLVLPLLAPAPSSAPKVRQSD---TWGPRSQWSPCSTRTCGGVSRFRPCYSORSD	59
QY	65	---SCEGRNIRYRTC SNVDCPPPEAGDFRAOQCSAHNDVKHHGQFYEWLVPNSDPNPCS	121
DB	60	GGSSCVGPARSHRSCRTESCPDGAERDFAEOCAFEAGPQRRYRWLPYSAP-NKCEL	118
QY	122	KCOAKGTYTVLVELAPKVLDTGTRCYTESLDMCISGLQTVGCDHQJLGSTVKEDNCGV	181
DB	119	NCIPKGFENFYKHBREAVYDTPCEPKRGDRCYVDGSCRVVCGDHELDSSKQEDKLC	178
QY	182	GSTCLRVGOYKSOLSATKSDDTVAIPYGSRRHRLVLKDPDHLIETKTLOG-----	234
DB	179	GTTCTYPVAGTFDAN-DLSRGYNQLLIVPMGATSLIDEAASRNFLAVNRVGEYYLN	237
QY	235	--TWGENSLSTGTFVLVNDSSVDFKPFDDKEILLRMAGPLTADFIVKIRNSGSADST	292
DB	238	WTIEAARALPAASTILHYERGAEGDLAPER--LHARGPTSEPLVIEL-LSQRENP	294
QY	293	FYQPLIH---RWRETDFFPCSATCGGGQQLLSAECYDLRSNRVADQVC---HYYPE	343
DB	295	YHLPRLRPSGFGSHWSGWSDCSAECGGGHQ-----SRLV---FCTIDHEAYD	340
QY	344	NI---KKP-KLQFCNLDPDPSADQYKIMPYDLYHLP-----RWEATPWTA	387
DB	341	HMCORQPRPADRRSCNLHPCPKRTSY-----LHRPGAWRLAGAQVCGNSWKGAP	395
QY	388	CSSSCGGGIGSRVCSVEEDIQGHVTSVEWKOMYTP-KMPIAQPCNIDFCPKWLAQ	446
DB	396	CSASCSCGGGSSRSYCTSSDGAGIQAVBEAEACAGLPKPPAQAACNLQRCAAWSP	455
QY	447	PCVTWTCOGULRYRWLCLDHRG--MHTGCSPKTKPHKECIVTPCYKPEKLPVEAK	504
DB	456	ECVSCVCGVGRKRSVTCRGERGSLHTAACSLDREPLTEPCY-----HEDCPL	507
QY	505	LPMPKQAOELEFEAAVSEEPSFTPEAWSACTVTCGVTOVLRVCOVLLSFSQSVAD	564
DB	508	QAWH-----VGTWGLCSKSCSGSTRRQVICA-----GPP	538
QY	565	DEC---EGPKPAPSORACYGPCSGEIPENPDITDGLFGGLDDELVDWEYEGTKCSE	621
DB	539	SHGCSLQHSKVPDVEPCNTQPC--HLQPQVSPMODVHTPASNPWPL-----	583
QY	622	SCGGGVQEAUVVSCINKOTREPAENICLVSRRRPQLLKSCNLDPCPARWEIG--	673
DB	584	----GFQESPASA-----APIDATPAVGLRAPRLQTOQSSRVLV---RWP	630
QY	674	KWSPCLSTCGVGLQTRDVFCSHLLSRENNETVILLADELCR-----	713
DB	631	LPWGP-----LSAEQVHNTHQQAQNE-----PSECRCQFQCCYDNDVAT	677
QY	714	----QPK-----PS-----TVQACNRFNCPWAWP-----	734
DB	678	EGCVGQPSHAYPVRCLLPSAHGSCADWAARWYFVASVGCNRF-----	732
QY	735	AQWQPCSTRCTGGGVQKREVLCVKORADGSFLELPETFCASAKPACQ-----	785
DB	733	ASQECMSQCQSLHG-----PRRPPGASGRSTHTDGGSSPAGQEIPSOHRTGA	787
QY	786	DDCPS-----EWLLSDWTE--CSTSCGGETOTRSAICRKLKT	821
DB	788	KMPWGGGLWRQDOQPCGPEAPHTQAFGEW---PWQELGSRAPGLGGARS-----	835
QY	822	GLSTVNVNSTLCPPLPSSSIRPCMCLTARCPGRPKTSHPSHTAAARKVYIOTROR	881
DB	836	-----PAPPPHSSSYRISLAGV-----EPSIVQA-----	859
QY	882	VVGGFAYLLPKTAVVLRRCFARRVRKPLTWEXDGOHLLISHTVTVAPPGYKIHRLK	941
DB	860	ALG-----QLVRJUSCDDTAPESQAWQKDGQIFSSDRH-RIQFGPSLIITHPLO	910

QY	942	AGYTC-SAGPAR--EHFVIKLIGGNKRLVARPLSPRSEEEVLAGRGGPKKALQTHKHQ	998
DB	911	AGTYSCTGSTRPGDSOKIQLRIIG--LCPHPIH-----HSH-	944
QY	999	NGIFSNCSAEKGLAANPGSRDDLVSLRLEOGGWPGEGLLASWEAODSAERNTTSEDP	1058
DB	945	---GAPQVLLHLPFTMTVTEORRLDDILGNLSOQPEELRDLYSKHLVAQLAEIFRSHLEHQT	1118
QY	1059	GAQVLLHLPFTMTVTEORRLDDILGNLSOQPEELRDLYSKHLVAQLAEIFRSHLEHQT	1118
DB	964	-----LSRPPQ-----	969
QY	1119	LLKPSERRTSPTVLSPHKHVSGFSSSLRTSSTGDAG--GGRSRPHRKPRTILTKISAAQQL	1176
DB	970	-----PRDPAQDFGQA-----GAAGPLGAIPSSHPOPANRLRLDQNO--	1006
QY	1177	SASEVVTHLGQTVVALASGTLVLLHCEAIGHPRPTISWARGNEVQFSDRILLQPDSDLQ	1236
DB	1007	-----PRVVDASPGQRIHMTCRAGFFPPPAIEWQDGFVS--SPRHQLQPDGSLV	1055
QY	1237	ILAPVEADYGFTTCNATNALGYDS--VSLAV-----TLAGKPLVKTRSMVTINTEKPAVTV	1290
DB	1056	ISKRAVEDGFGFTCYAFNGQDRDQRVQLRVLCGLIISGLP-----FTVTV	1101
QY	1291	DTGSTIKTVQGVNVTINCQVAGPDEAVETWFRNKSGLSP-----HHLH--EGSLLLTN	1342
DB	1102	PGEDTARLL-----CVVAG--ESYNIRMSRN---GLPQVADGHRVHQSPDGTLLIYN	1148
QY	1343	VSSSQGLYSCRAANLHGLTESTQLLILOPPQVPT	1378
DB	1149	LRADEGSVTCSNAYQGSQAVSRSTEYKVVSPP--APT	1182

RESULT 11

O76840 PRELIMINARY; PRT; 2167 AA.

ID O76840

AC O76840; Q22911;

DT 01-NOV-1998 (JREMBLrel. 08, Created)

DT 01-NOV-1998 (JREMBLrel. 08, Last sequence update)

DT 01-DEC-2001 (JREMBLrel. 19, Last annotation update)

DE C37C3.6 PROTEIN.

GN C37C3.6.

OS Caenorhabditis elegans.

OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;

OC Rhabditidae; Platyhelminthes; Caenorhabditis.

OX NCBI_TaxID=6239;

RP [1]

RN SEQUENCE FROM N.A.

RA STRAIN=BRISTOL N2;

RA Geisel C., Bradshaw H.;

RT "The sequence of C. elegans cosmid C37C3.6";

RL Submitted (JUL-1996) to the EMBL/GenBank/DBJ databases.

CC -!- ALTERNATIVE PRODUCTS: TWO FORMS (A AND B) MAY BE PRODUCED BY

CC ALTERNATIVE SPLICING OF THE SAME GENE. THE SEQUENCE SHOWN IS THAT

CC OF FORM B.

DR EMBL: U64857; AAC25868.1; -

DR EMBL: U64857; AAC25867.1; -

DR HSSP: P00981; IDPK.

DR InterPro: IPR003598; Ig_C2.

DR InterPro: IPR003006; Ig_MHC.

DR InterPro: IPR002223; Kunitz_BPTI.

DR InterPro: IPR000884; TSPL.

DR Pfam: PF00047; Ig; 1.

DR Pfam: PF00014; Kunitz_BPTI; 11.

DR Pfam: PF00090; Tsp_1; 6.

DR PRINTS: PR00759; BASICPTASE.

DR SMART: SM00408; IGC2; 1.

DR SMART: SM00131; KU; 11.

DR SMART: SM00209; TSPL; 7.

DR PROSITE: PS00280; BPTI_KUNITZ_1; 10.

DR PROSITE: PS02079; BPTI_KUNITZ_2; 11.

DR PROSITE: PS50092; TSPL; 4.

DT	01-MAR-2001	(TrEMBLrel. 16, Created)			
DT	01-MAR-2001	(TrEMBLrel. 16, Last sequence update)			
DT	01-DEC-2001	(TrEMBLrel. 19, Last annotation update)			
DE	INTRACELLULAR MATRIX PROTEIN PAPILIN PRECURSOR.				
GN	PPN.				
OS	Drosophila melanogaster (Fruit fly).				
OC	Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;				
OC	Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;				
OC	Ephydroidea; Drosophilidae; Drosophila.				
OX	NCBI_Taxid=7227;				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RC	STRAIN=DP CN BW;				
RC	MEDLINE=20530499; PubMed=11076767;				
RA	Kramerova I.A., Kawaguchi N., Nelson R.E., Fessler L.I., Chen Y.,				
RA	Kramerov A.A., Kusche-Gullberg M., Kramer J.M., Ackley B.D.,				
RA	Sieron A.L., Prockop D.J., Fessler J.H.;				
RT	"Papilin in development; a pericellular protein with a homology to the				
RT	ADAMTS metalloproteinases";				
RL	Development 127:5475-5485(2000).				
DR	EMBL; AF205357; AAG37995.1; -.				
DR	HSSP; P12111; 2KNT.				
DR	FlyBase; FBgn0003137; Ppn.				
DR	InterPro; IPR003599; Ig.				
DR	InterPro; IPR003598; Ig_c2.				
DR	InterPro; IPR003600; Ig_like.				
DR	InterPro; IPR002223; Kunitz_BPTI.				
DR	InterPro; IPR000884; TSP1.				
DR	InterPro; IPR002221; WAP.				
DR	Pfam; PF00014; Kunitz_BPTI; 3.				
DR	PRINTS; PR00003; 4DISULPHCORE.				
DR	PRINTS; PR00759; BASICPTASE.				
DR	SMART; SM00409; IG; 3.				
DR	SMART; SM00408; IGC2; 2.				
DR	SMART; SM00410; IG_like; 1.				
DR	SMART; SM00131; KU; 3.				
DR	SMART; SM00209; TSP1; 7.				
DR	SMART; SM00217; WAP; 1.				
DR	PROSITE; PS50279; BPTI_KUNITZ_2; 3.				
DR	PROSITE; PS50092; TSP1; 3.				
KW	Matrix protein; Serine protease inhibitor; Signal.				
FT	SIGNAL 1 26 POTENTIAL.				
SQ	SEQUENCE 2174 AA; 231936 MW; 038F07952623120 CRC64;				
Query Match 9.2%; Score 882; DB 5; Length 2174;					
Best Local Similarity 28.0%; Pred. No. 5e-55;					
Matches 224; Conservative 94; Mismatches 282; Indels 200; Gaps					
QY	37	DANGPW---SECSRTGGGASYSLRCL-----SKSCGGRNIRYRTCSNVDCPEAGDF 88			
DB	58	DEWTPWSSPSDCSNTCGGVSYQTRCLRDRDRGRAVCGSGRRYFSCNTQDCPEESDF 117			
QY	89	RAQCCSAHNDVKHHGQFYEWLPSVNDPDPNCSLKQCAAGTTLVLEAPKVLDDGTRCYTES 148			
DB	118	RAQCCSFDRQQDFGVFEYEWYPTNAP-NPELCNMPKGERFYEQREKVVDSGTFRCDKD 176			
QY	149	LDMCISGLQCIQVCDHQLGSTVKEDNCVCGDGTCLRIVRQYKQSLSATKSDTVAI 208			
DB	177	LDVCVNGECPVGCMDMLGSDAKCKRCGGDGGSTCKTRITITTKLAFGYND-LILL 235			
QY	209	PYGSRRHRLVILKGDHLYLETKLQG--TKGENSLSTGTFLVDNSSVDFOK---FPD 261			
DB	236	PEGATNRIIEETWPSNYYLACRNHSGHYLYNGWRIDIPRPFMFANSWNTQRPDMGFAA 295			
QY	262	KEILRMAGPLTAD-FIVKIRNSGSADSTVQFIYQFI-----IHRWRETDFPCCATCG 314			
DB	296	PQQLTCSGPISLSIFVIMLVQEKNISLDVEYSIPESLSHSQDQDTHWTHQFNACSCSG 355			
QY	315	GGVQLTSAECYDILRSNV-----VADQYCHYPENIKPKP-KLQCNLDPCPADGY 365			
DB	356	GGSQNRKRVTC---NNRITILAEVNPSLDQ-----KSKPVEEQACGTEPC----- 396			

```

QY 366 KQIMPYDLYHPLPWEATPTWACSSCGG-GIQSRVAVSCVDEEDIQGHVTSVEENKCM--Y 422
Db 397 -----APHWEGESKSGKSGDGFONRSITCRITSSGSEHTVEEDAVCLKEV 445
QY 423 TPKMPIAOPCN--TFDCPKWLAQWSPCTVTCGGGLRVYRVLCIDHRCMHGTGCGSPKTKP 480
Db 446 GNKPATYQECNRDYNKCHYHLGPTWPCDKLGGDKGQTKRKYTCFTIENGHK----- 496
QY 481 HIKBECIVPTPCYKPKKEKLPYEAKLPMWFKQAEELEGAASVEEPSFIPEANSACTVTCGV 540
Db 497 ----- 496
QY 541 GTQVRIYRCVLLSFQSQVADLPIDECBGPXPSQACRACYPAGPCSGEITPEFNPDETDLFG 600
Db 497 -----LPEDCEVEKPEKTEKSLLPCEG----- 522
QY 601 GLODFDELYDWEYEGFTKCSCEGGGVQEAUVVCLNKOTREPAENLCVTSRRPPQLLKS 660
Db 523 -----VDMIIISWSSCN-ACQONTETRAICGNKEGKYVPEE-FC--RPEVPTLSRP 570
QY 661 CNLDPCEARWEIGKWSLTCGGVGLQTRDVFCSHLLSREMNETVILADE--LC-RQPKP 717
Db 571 CKSPKCEAOWFSSEWSKCSAPCGKGVKSRVICGEF----DGKVTVPADDDSKCNKETKP 626
QY 718 STVOACNRN--CPANYPACWQPCSRFCGGGVOKREVICKQRMADGSFELPETFCSAS 775
Db 527 ESEDCGEEKVCVGEWFTGPMGCKSPKCGGGERVREVLG---LSNGT-----KSVNCDDE 679
QY 776 K-PACQOACKKDDCPSEWLL 794
Db 680 KVEPLSEKCNSEACTEDEL 699

RESULT 13
Q9VAV4
ID Q9VAV4 PRELIMINARY; PRT; 3060 AA.
AC Q9VAV4; Q9VAV3;
DT 01-MAY-2000 (TEMBLrel. 13, Created)
DT 01-MAR-2001 (TEMBLrel. 16, Last sequence update)
DE 01-DEC-2001 (TEMBLrel. 19, Last annotation update)
DE CG1540 PROTEIN.
GN PPN OR CG1540 OR CG18436.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
ON NCBI_TaxID=7227;
RX STRAIN=BERKELEY;
RC MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazer R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C.A., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Ballou R.M., Basu A., Baxendale J., Andrews-Pfannkoch C., Baldwin D.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Beasley E.M.,
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brotier P.,
RA Burtis K.C., Busam D.A., Butler J., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doul L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Foster C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Keichum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,

```

```

RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA Palazolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svirskas R., Tector C., Turner E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weissbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of Drosophila melanogaster."
EL Science 287:2185-2195(2000).
CC 1- ALTERNATIVE PRODUCTS: 2 ISOFORMS; A LONG FORM (SHOWN HERE) AND A
CC SHORT FORM; ARE PRODUCED BY ALTERNATIVE SPLICING.
DR EMBL; AE003765; AAF56794.2; -
DR HSSP; P12111; 2KNT.
DR FlyBase; FBgn0003137; Ppn.
DR InterPro; IPR0003561; EGF-like.
DR InterPro; IPR003598; Ig_C2.
DR InterPro; IPR003006; Ig_MHC.
DR InterPro; IPR002223; Kunitz_BPTI.
DR InterPro; IPR000884; TSPI.
DR InterPro; IPR002221; WAP.
DR Pfam; PF00047; Ig_3.
DR Pfam; PF00014; Kunitz_BPTI; 12.
DR Pfam; PF00090; tsp_1; 5.
DR Pfam; PF00095; wap; 1.
DR PRINTS; PR00759; BASICPTASE.
DR SMART; SM00408; IGC2; 3.
DR SMART; SM00131; KU; 12.
DR SMART; SM00209; TSPI; 7.
DR SMART; SM00217; WAP; 1.
DR PROSITE; PS00317; 4_DISULFIDE_CORE; 1.
DR PROSITE; PS00280; BPTI_KUNITZ_1; 11.
DR PROSITE; PS00279; BPTI_KUNITZ_2; 12.
DR PROSITE; PS00022; EGF_1; UNKNOWN_1.
DR PROSITE; PS00092; TSPI; 3.
KW Alternative splicing; Immunoglobulin domain;
KW Serine protease inhibitor.
FT VARSPPLIC 2803 2803 L-> SVTPV (IN SHORT ISOFORM).
FT VARSPPLIC 2844 2854 FNFKTMSDGI-> VASPLHPNAV (IN SHORT
FT VARSPPLIC 2855 3060 ISOFORM).
FT VARSPPLIC 3060 AA; 331579 MW; ACA31D3EE558C7C0 CRG64;
SQ SEQUENCE

Query Match 9.28; Score 882; DB 5; Length 3060;
Best Local Similarity 28.08; Pred. No. 8.6e-55;
Matches 224; Conservative 94; Mismatches 282; Indels 200; Gaps 29;

QY 37 DAWGPW---SECSRTCGGASYSILRRCL-----SSKSCGGRNIRYTCNSVDCPPEACDF 88
Db 18 DEWPWSPSDCSCTCGGVSQYQRECLRRDDRGEAVCSGSRIFSCNTODCPPEESDF 77
QY 89 RAQCSAHNDVKHGGQFYEWLPVSNPDNPNCSLAKCOAGTTLVVVELAPKVLDPGRCYTES 148
Db 78 RAQCSRFDRQQDFGVFEWVPVYNAP-NPCELNCMPKGERFYRQREKVDGTRCNDKD 136
QY 149 LDMCISGLCIVGCDHQLGSTVVEDNCGVNGDSTCLRVGQYKSLSTKSDTVAI 208
Db 137 LDVCVNECPVPCGDMGMGDAKDKCRGGGDKCTKTIRNTITTKDLAPGYND-LLLL 195
QY 209 PYGSRHRLVLKGDHLYLETKTLQG---TKGENSLSTGTFLVDNSVDFQK----FPD 261
Db 196 PEGATNRIETVPSNLYACRNHSHGYINGDRIIDFPRPMFFANSWNVYQRKPMGF 255
QY 262 KEITRMAGPLTAD-FIVKIRNSGSADSTVQIFVQPI-----IHRWRETDFFPCSATCG 314

```

```
Db 256 PDQLTSGPISESLFIVMLVQEKNISLDYEISIPESLSHSQDTHWTHHQFNACSASCG 315
QY 315 GGQLTLSACYDLRSNRV-----VADQYCHYYPENTKPKP-KLQECNLDPASPADGY 365
Db 316 GGSQSRKVIC-----NKRITLAEVNPISLQDQ-----KSKPVEQACTEPC----- 356
QY 366 KQIMPYDLHPLPRWEATPWTACSSCGG-GIQSRAVSCVEEDIOGHVTSVEEWKCM--Y 422
Db 357 -----APHWGEWSKSKSCGGSGDFQNRISITCERISSGHEHTVEEDAVCLKEV 405
QY 423 TPKNPIAQPCN--IFDCKPKLADQWSPCTVTCGGGURVYRVVLCIDHRCMTGGSPKTKP 480
Db 406 GNPATKQECNRDVKNCPKYHLGFWPTPCDKLGDGKGTRKVTCTFIEBNGHK----- 456
QY 481 HKEECIVPTPCYKPKKLEKLPWFKAQLEEGAAVSEBPSFIPEAWSACTVTCGV 540
Db 457 ----- 456
QY 541 GTQVRIVRQVLLSFQSVADLPIDECEGPKPASQACRACYPAGSCGEIPEFNPDETGLFG 600
Db 457 -----RV-----LPEDCVBEKPEKESCLLTPCBG----- 482
QY 601 GLQDFDELYDWEYEGFTKCSSESGGGVQEAIVVSLNKTQREAPENLCVTSRRPPQLLKS 660
Db 483 -----VDWIIISQWSGCN-ACQONTETRTAICGNKEGVYEE-FC--EPEVPTLSRP 530
QY 661 CNLDPCEARWEIGKWSPCSLTCGVLQTRDVFCSHLLSRENNETVILADE--LC-RPKP 717
Db 531 CKSPKCAQWFSSEWSKCSAPCGKGVASRIVICGEF-----DGKIVTADDDSKCNKETKP 586
QY 718 STVQACNRFN--CPPAWYPAQWPCSTCGGGVQKREVLCKQRMADGSFLELPFTFCSAS 775
Db 587 ESEQDCGEKEKVCPEWFTGFWGKSCPKCGGGERVREVLG---LSNCT---KSVNCDEE 639
QY 776 K-PACQOACKKDCPCSEWLL 794
Db 640 KVEPLSEKCNSEACTEDEL 659

RESULT 14
ID O44938 PRELIMINARY; PRT; 1572 AA.
AC O44938;
DT 01-JUN-1998 (Tremblrel. 06, Created)
DT 01-JUN-2001 (Tremblrel. 17, Last sequence update)
DT 01-DEC-2001 (Tremblrel. 19, Last annotation update)
DE THROMBOSPONDIN.
GN THRI.
OS Haemochus contortus.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Strongylida;
OC Trichostrongyloidea; Haemonchidae; Haemonchidae; Haemonchidae.
OX NCBI_TaxID=6289;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-MOREDUN;
RA Skuce P.J., Newlands G.F.J., Stewart M., Pettit D., Smith D.,
RA Knox D.P.;
RT "Cloning and characterization of thrombospondin, a novel multidomain
RT glycoprotein associated with the gut of Haemonchus contortus.";
RL Submitted (MAR-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF043121; AAB99830.2; -
DR HSSP; P05067; ICA0.
DR InterPro; IPR000561; EGF-like.
DR InterPro; IPR002223; Kunitz_BPTI.
DR InterPro; IPR000884; TSP1.
DR Pfam; PF00014; Kunitz_BPTI; 12.
DR Pfam; PF00090; Tsp1_1; 6.
DR PRINTS; PR00759; BASICPTASE.
DR SMART; SM00001; EGF_like; 1.
DR SMART; SM00131; KU; 6.
DR SMART; SM00209; TSP1; 7.
DR PROSITE; PS02079; BPTI_KUNITZ_2; 6.
DR PROSITE; PS00092; TSP1; 2.
```

```
KW Serine protease inhibitor.
SQ SEQUENCE 1572 AA; 171871 MW; 2260B30DC2F903EC CRC64;

Query Match 9.08; Score 865; DB 5; Length 1572;
Best local Similarity 28.2%; Pred. No. 5.3e-54;
Matches 224; Conservative 102; Mismatches 279; Indels 188; Gaps 30;

QY 24 RTARSEEDRDLWDAPGW--SECSRTCCGGGASYSLSRCLSSKSGEGRNIRYRTCSNVDC 81
Db 66 KTGSES-----GPGWFWPEQCSRTCCGGGVQTEKQC--PGDCTGPSVYVSCNLEPC 117
QY 82 PPEAGDFRAQOCSAHNDVKHGFYEWLPVSNPDNPSLCKQAKGTTLVVVELAPKVLGD 141
Db 118 -PECTDFRAEQCAAHNDPDDGQYHKWIPYKG--KNKCELLCKPENGFFYKWDFTVVDG 174
QY 142 TRCYTESLDMCISLQIVGCDHQLGSIVKEDNGVCNGDSTCLRVGOY--KSOLSATK 200
Db 175 TKDSKSGDDICVDGVCPLGCDGKLSALKLDKCGVCDGSGQCKTIEGSEDERNLSPGY 234
QY 201 SDDTVVAIPYGSRRHRIWLVKGPDPHLYLETKTLOCTKGNSLSSTGTFLVDNSSVDFQK-- 258
Db 235 HD--VMRIPAGATAIRIEEARPSSNNLAMKN-----SSDYIFLNGNSMIQVEKD 282
QY 259 -----FPD--KEILLRMAGPLTADFTIVKTR-NSGSADSTVQVFIYQPII-----HRWRE 303
Db 283 ELNGVHFHEYDDGKPERITAKGPLLEEYVVSVLIRKGNRDVSIKYEFSVPIYDEVDDMYKP 342
QY 304 TDFPSCATCGGGYOLTSACVYDLRSNRVADQYCHYYPENIKPKKLOECNLDPCCASD 363
Db 343 GEWSACSVTCKGQGTTPYCIETKTQARVDDQLC--DDANSTKPEFEK-----PCETVD 395
QY 364 GYKQIMPYDLHPLPRWEATPWTACSSCGG-GGIQSRVSCVEEDIOGHVTSVEEWKCMY 422
Db 396 QCAE-----WFKGEWEPSCQTCGQGEQYRVVYCHQVFANGRRITVDDGNC-- 441
QY 423 TPKNP-IAQPCNIFDCKPKLADQWSPCTVTCGGGURVYRVVLCIDHRCMTGGSPKTKPH 481
Db 442 TAERPVRQVQCNRFACPEWQAGPWSACSEKCGDAFYQYRSVTC----- 483
QY 482 IKECIVPTPCYKPKKLEKLPWFKAQLEEGAAVSEBPSFIPEAWSACTVTCGV 541
Db 484 -----RSEKEGEGLK----- 494
QY 542 TQVRIVRQVLLSFQSVADLPIDECEGPKP-ASQACYPAGSCGEIPEFNPDETGLFG 600
Db 495 -----LPAAECSSSENTLESQSCNLGPT----- 518
QY 601 GLQDFDELYDWEYEGFTKCSSESGGGVQEAIVVSLNKTQRE-PAENLCVTSRRPPQLLK 659
Db 519 GLKFFTS--EWKL-----CSK-CNDTEETREYTKDSQGRAYPLEKCLTDDKEKEIPVDTR 570
QY 660 SC-NLDPCEARWEIGKWSPCSLTCGVLQTRDVFCSHLLSRENNETVILADELCKQPKPS 718
Db 571 ACATQPCPNYEWTSQNSKSCSTECGHHKTRVICA---IHEEGDVTVVDEALCOGEKPE 627
QY 719 TVQAC-NRENCPPAPYPAQWPCSTCGGGVQKREVLCKQRMADGSFLELPFTFCSASKP 777
Db 628 VKMNCNTNEKCTGTWYSGPWSPCSVCEGCGKQERVAVC----LNYDKKPPVPEWCDESEMP 683
QY 778 ACQOACKKDDCPES 790
Db 684 NLIQDCNTDECPES 696

RESULT 15
ID Q908G8 PRELIMINARY; PRT; 3198 AA.
AC Q908G8;
DT 01-MAY-2000 (Tremblrel. 13, Created)
DT 01-MAY-2000 (Tremblrel. 13, Last sequence update)
DT 01-DEC-2001 (Tremblrel. 19, Last annotation update)
DE LACUNIN PRECURSOR.
```


GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 24, 2002, 02:20:32 ; Search time 61.07 Seconds
(without alignments)
3204.716 Million cell updates/sec

Title: US-10-044-807-2
Perfect score: 9588
Sequence: 1 MECCRAITPGTLLFLAFLL.....LKLCOLFSSKCCGTCGKA 1762

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 747574 seqs, 111073796 residues

Total number of hits satisfying chosen parameters: 747574

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A_Geneseq_032802.*
1: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1980.DAT.*
2: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1981.DAT.*
3: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1982.DAT.*
4: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1983.DAT.*
5: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1984.DAT.*
6: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1985.DAT.*
7: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1986.DAT.*
8: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1987.DAT.*
9: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1988.DAT.*
10: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1989.DAT.*
11: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1990.DAT.*
12: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1991.DAT.*
13: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1992.DAT.*
14: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1993.DAT.*
15: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1994.DAT.*
16: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1995.DAT.*
17: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1996.DAT.*
18: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1997.DAT.*
19: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1998.DAT.*
20: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1999.DAT.*
21: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA2000.DAT.*
22: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA2001.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	9466.5	98.7	1745	22 AAB90551	Human secreted pro
2	9464.5	98.7	1766	22 ABB10246	Human CDNA SEQ ID
3	4797.5	50.0	874	21 AAB01431	Human TANGO 224 (f
4	3956.5	41.3	1691	22 AAE07863	Novel human protei
5	3953.5	41.2	1691	22 AAE10609	Human novel KIAA12
6	3762.5	39.2	1617	22 AAE07870	Novel human protei
7	3565	37.2	643	21 AAY87347	Human signal pepti
8	3423	35.7	645	22 ABB10454	Human CDNA SEQ ID
9	2917	30.4	525	20 AAY41729	Human PRO1071 prot
10	2917	30.4	525	22 AAU29241	Human PRO1071 (UNQ
11	2907	30.3	525	21 AAB44285	Human PRO1071 (UNQ

12	2907	30.3	525	22 AAB72290	Human ADAMTS-R1 am
13	2809.5	29.3	845	22 AAE07868	Novel human protei
14	2615.5	27.3	771	22 AAE07869	Novel human protei
15	2588	27.0	514	22 AAB30586	Human secreted pro
16	2567.5	26.8	1281	22 AAE10608	Human novel KIAA12
17	2567.5	26.8	1281	22 AAE10610	Human novel KIAA12
18	2425.5	25.3	724	22 AAE07866	Novel human protei
19	2283.5	23.8	417	22 AAG78679	Human thrombotic p
20	2282.5	23.8	926	22 AAB83370	NOV14 protein sequ
21	2231.5	23.3	650	22 AAE07867	Novel human protei
22	2198	22.9	428	21 AAB41561	Human ORFX ORF1325
23	1981	20.7	480	21 AAB01425	Human TANGO 224.
24	1930	20.1	368	22 ABB11218	Human BPTI/Kunitz
25	1505	15.7	292	22 AAB30632	Human secreted pro
26	1438	15.0	446	22 AAE07864	Novel human protei
27	1388	14.5	262	22 AAB30587	Human secreted pro
28	1353	14.1	886	21 AAB40500	Human ORFX ORF264
29	1244	13.0	372	22 AAE07865	Novel human protei
30	1145.5	11.9	2150	21 AAY33898	Amino acid sequenc
31	1145.5	11.9	2165	22 AAB30617	Human secreted pro
32	1111	11.6	281	22 ABB11315	Human HSP-124 pro
33	985.5	10.3	1934	22 AAB72301	Human ADAMTS-9 alt
34	972.5	10.1	1882	22 AAB72286	Human ADAMTS-9 ami
35	961	10.0	180	20 AAY35899	Extended human sec
36	947	9.9	337	22 ABG27865	Novel human diagno
37	886	9.2	207	22 AAM25476	Human protein sequ
38	884.5	9.2	1784	21 ABA11379	Human ORFX ORF1143
39	882	9.2	2858	22 ABB71150	Drosophila melanog
40	882	9.2	3060	22 ABB58084	Drosophila melanog
41	870	9.1	169	22 ABG27864	Novel human diagno
42	843	8.8	1686	22 AAB74944	Human ADAM type me
43	831	8.7	1686	22 AAE00913	Human 27875 ADAM-T
44	831	8.7	1690	22 AAB86949	Human metalloprote
45	830	8.7	1686	22 AAE00934	Human 27875 ADAM-T

ALIGNMENTS

RESULT 1
AAB90551
ID AAB90551 standard; Protein; 1745 AA.
XX AC
XX AAB90551;
XX 01-JUN-2001 (first entry)
XX Human secreted protein, SEQ ID NO: 89.
XX DE
XX Human; secreted protein; immunomodulatory; antisclerotic;
KW dermatological; antiinflammatory; anti-HIV; cytostatic; cardiant;
KW vascular; anti-angiogenic; ophthalmological; neuroprotectant;
KW neotrophic; anticonvulsant; antialzheimers; antiparkinsonian;
KW antimicrobial; vulnerary; vaccine; gene therapy; cancer;
KW protein coordinate data; infection.
XX OS
XX Homo sapiens.
XX WO200121658-A1. Requested Here
XX 29-MAR-2001.
XX 22-SEP-2000; 2000WO-US26013.
XX 24-SEP-1999; 99US-0155709.
XX (HUMA-) HUMAN GENOME SCI INC.
XX Ni J. Baker KP, Birse CE, Ebner R, Fiscella M, Komatsoulis GA;
PI Lafleur DW, Moore PA, Olsen HS, Rosen CA, Ruben SA, Soppet DR;
PI Young PE, Wei P, Florence KA;
XX WPI; 2001-235311/24.

DR	N-PSDB; AAF97891.	601	GLQDFDELYDWEYEGFTKCSGCGGVOEAVVCSLQKOTREPAEENLCVTRRRPQLLKS	660
XX	Nucleic acids encoding 32 human secreted polypeptides, useful for			
PT	preventing, diagnosing and/or treating e.g. cancers, Parkinson's	584	glqdfdelydweyegftkcsescgvgqavvsclnkqtrepaeenlcvtstirrpqliks	643
PT	disease and diabetic retinopathy -			
XX	Claim 11; Fig 4; 890pp; English.	661	CNLDPCPARWEIGKWSPCSLTCGVLQTRDVFCSHLLSRENNETVILLADELCROPKSTV	720
PS				
XX	The present sequence is one of 32 novel human secreted polypeptides. The	644	cnldpcparweigkwspscltcgvgltqtrdvfcshllsremnetvilladelcrpdkpstv	703
PS	nucleic acid molecules and polypeptides may be used in the prevention,			
CC	diagnosis and treatment of diseases such as immune disorders (e.g.	721	QACNRNCPAPWYPAQWQPCSRCTCGGVOKEVLCQRMADGSELELPETFCSSKAPACQ	780
CC	multiple sclerosis, systemic lupus erythematosus and human			
CC	immuno-deficiency virus (HIV) infections), hyperproliferative disorders	704	qacnrficppawypaqwqpcsrctcgvgqkrevickqrmadgsflelpeticssaskpacq	763
CC	(e.g. cancers and Gaucher's disease), cardiovascular diseases			
CC	(e.g. Sclimtar syndrome, Chaga's cardiomyopathy and coronary	781	QACKKDCPSWLLSDWTFECTSCGEGQTRSAICRKMKTGLSTVNVNSTICPPLPSSSS	840
CC	arteriosclerosis), androgenic disorders (e.g. corneal graft			
CC	neovascularisation and diabetic retinopathy), neurological disorders	764	qackkddcspewllsdwtectscgegltqtrsaicrkmktglstvnvsticpplpfssss	823
CC	(e.g. Huntington's chorea, Alzheimer's disease and Parkinson's disease),			
CC	infectious diseases and/or for promoting wound healing, regeneration	841	IRPCLATCARPGSPSTKHSPIAARVYIQTRORLHFVVGFAVLLPDKTAVVLRCP	900
CC	and/or chemotaxis. The nucleic acid molecules may be used to produce the			
CC	secreted polypeptides. They may also be used as DNA probes in diagnostic	824	irpclatcarpgpstkhspshiaarkvyiqtrqrkhrfvvgfayllpdktavvircp	883
CC	assays to detect and quantitate the presence of similar nucleic acid			
CC	sequences in samples. The polypeptides may be used as antigens in the	901	ARRVRKPLITWEKDGQHLISSTHTVAPFGYLKTHRLKPSDAGVYTCISAGPAREHFVTKL	960
CC	production of antibodies and in assays to identify modulators of			
CC	their expression and activity.	884	arrvrkplitwekdqhlissthvfvapfgykhrlkpsdagvytcsagparehfvkl	943
XX	Sequence 1745 AA;			
SQ	Query Match 98.7%; Score 9466.5; DB 22; Length 1745;	961	IGNRKXIVARPLSPRSEEVLAGRGGPKALQTHKHONGIFSNCSKAERGLAANPQSR	1020
	Best Local Similarity 99.0%; Pred. No. 0;			
	Matches 1744; Conservative 1; Mismatches 0; Indels 17; Gaps 1;	944	iggnrkivarpplsprseevlagrkpkealqthkhqngifsnsgskaerkglaanpqs	1003
QY	1 MECCRRATPGTLLFLAFLLLSSRTARSEEDRGLDWAQGPWSCSRCTCGGASYSLRRC	1021	YDDLVRLLLEOGWPPELLASWEAODSAERNTTSEEDPAEQVLLHLPFTWTEQRRLDD	1080
DB	1 meccrratpgtllflaflllssrtarseedrdglwagpwwscsrctcgggasyslrcc			
QY	61 LSSKSCBGRNIRYTCNVDCPPBAGDFAQQSAHNDVXHHGQFYEWLPSVNDPNDPCS	1004	yddlvrlllegggwpgeallaweaqdsarnntseedpaaeqvllhlpftwteqrlldd	1063
DB	61 lsskscegnlrytcsnvdncppeagdfraqqcsahndvkhgqfyewlpvsndpndpcs			
QY	121 LKQAKCTTIVLAPKVLGTRCYTSLDWCISGLCQIVGCCDHLGTVKEDNCGVNG	1081	ILGNLSQPEELRDLYSKHLVAQAEIFRSHLEHQDTLLKPSERTSPVTLSPRKHVSG	1140
DB	121 lkqakgttllvvelapkvlgdtrcytesldmcisglcqivgcdhqlgstvkedncgvng			
QY	181 DGSTCRIVRGYKSQLSATKSDDPVVAIPYXGSHRILVKGPDHLYLETKLQGTGENS	1064	ilgnlsqpeelrldyskhlvqlaqlaeifrshehqdtllkpsertspvtilsphkhvsg	1123
DB	181 dgstcrivrgyksqlsatksdttvvaipygshrilrvlkgpdhlyletklqgtgens			
QY	241 LSSTGTFLVONSVDFOKFPDKKEILRMAGPLTADFTVIRNSGSADSTVQPIFYQPIIHR	1141	FSSSLRYSSTGAGGSRPRHKKPTILRKISAAQQLSASEVTHLQGVVALASGTLVILL	1200
DB	241 lsstgtflvdnssvdfkfpdkellrmagpltdftvirknsgsadstvgfivqpiih			
QY	301 WRETFDFPCSATCGGYQLTSAECYDLRSNRVVADQCHYYPENIKPKPLQECNLDPCP	1124	fssslrstsstgaggsrrphrkptilrkisaaqqlsasevthlqgtvalasgtlsvll	1183
DB	301 wrecdfpcsatcggyqyltsaecydlrsnrvvadqchyyypenikpkplqecnlpcp			
QY	361 ASDGYKQIMPYDLHPLFRWEATPTWTACSSSCGGGIGSRVSCVEEDIQGHVTSVEEWKC	1201	HCEAIGHPRPTISWARNGEEVQFSDRIQLQPDSDLSQIILAPVEADVGFYTCNATNALGYS	1260
DB	361 a-----rweatptacsccgggiqsravscveediqghvtsveewkc			
QY	421 MYTPKMPIAQPCNIFDCKPKLAQSWSPCTVTCGGGLRYRVVLCIDHRGMHTGGSPKTKP	1184	hceaiighprptiswarngeevqfsdrillqpdslsqilapveadvgfycnataalgysd	1243
DB	404 mytpkmpiaqpcnifdcpkwlaqewspctvtcgqglryrvvclidhrgmhtggcspktp			
QY	481 HIKEECIVPTCYKPKELPVEAKLPWFKOAELEEGAASVEEP-SFIPKWSACTVTGCV	1261	VSIATVLAGKPLVKTSMRTVINTEKPAVTVDIGSTIKTVQGVNVTINCOVAGVPEAEYTW	1320
DB	464 hieecivptcykpkelpveaklpwfkaqealeegaasveepsfipkwsactvtgcv			
QY	541 GTQVRIVRQVLLSFSQVADLPIDECEGPKPASQRACYAGPCSGEITPEFNPDETDGLFG	1244	vsiavtlagkplvktstmrvtintekpavtdigstiktvggvnvtincqvagvpeaevtw	1303
DB	524 gtqvrivrqvlisfsqsvadlpidecegpkpasqracyagpcsggeipefnpdetdglfg			
		1321	FRNKSGLSPHLLHSGSLLLLNVSSDGLYSCRAANLHGETESTQILLILDPPQVPTQL	1380
		1304	frnkskigsphllhsgsllltnvssddgylsraanlhgeltetqllildppqvptql	1363
		1381	EDIRALLAATGPNLPSVLTSPGLQVLDPGNSALLGCPKIGHPPVNTWTFHGGQPIVTA	1440
		1364	edirallaatgpnlpvltspglqvlidpugnallgcpkighppvntwtfhggqpipta	1423
		1441	TGLTHHILAAQGILOVANLSGGSGQGEFSCLAQNEAGVLMQKASLVIQYWNVSVDRLATCS	1500
		1424	tgllthhilaagqilqvanlsgsgqgefscilaqneagvlmqkaslvigqywwwsvdrlatcs	1483
		1501	ASCGNRGVOQPRLCILNSTEVNPAHCAGKVPVAPQIACNPRDCPSRMWVTSWACTHRS	1560
		1484	ascgnrgvoqprrcllnstevnphacagkvpvqpqlacarrddcpsrmwvtswacthrs	1543
		1561	CGGGVQTRRVTCQKLLKASGISTPVSNDMCTQYAKRFPVDTQACNQQLCVWEAFSSWGQNG	1620
		1544	cgggvqtrrvtcqkllkasgistpvsndmctqykrpvdtaqcnqlcvewafsswgqng	1603
		1621	PCIGPHILAVQHRQVFCQTRDGTITLPSQCSALPRPVSTQNCWSEACSVHWRVSLTCLTA	1680
		1604	pcigphilavqhrvfcqtrdgtitlpsqcsalprpvstqncwseacsvhwrvsltlcta	1663
		1681	TCGNYGFSRRRVECVHARTNKAVPEHLCSWGPBPANQRCNITPCENMECRDTRTYCEKV	1740

|||||
Db 1664 tcgnygfsrrvecvhartnkavpehlcswgprpanwgrcnitpcenmecrdttrycekv 1723
QY 1741 KQLKLCQLSQFKSRCCGCGKA 1762
Db 1724 kqikicqlsqfskrcgctgka 1745
RESULT 2
ABB10246
ID ABB10246 standard; Protein; 1766 AA.
XX ABB10246;
XX
DT 10-JAN-2002 (first entry)
XX
DE Human CDNA SEQ ID NO: 554.
XX
XX Human; gene therapy; neural disorder; immune system disorder;
KW muscular disorder; reproductive disorder; gastrointestinal disorder;
KW pulmonary disorder; cardiovascular disorder; renal disorder;
KW proliferative disorder; inflammation.
XX
OS Homo sapiens.
XX
XX WO200154474-A2. *Aggregated have*
XX 02-AUG-2001.
XX
XX 17-JAN-2001; 2001WO-US01349.
XX
XX 31-JAN-2000; 2000US-179065P.
PR 04-FEB-2000; 2000US-180628P.
PR 24-FEB-2000; 2000US-184664P.
PR 02-MAR-2000; 2000US-186350P.
PR 16-MAR-2000; 2000US-189874P.
PR 17-MAR-2000; 2000US-190076P.
PR 18-APR-2000; 2000US-198123P.
PR 19-MAY-2000; 2000US-205515P.
PR 07-JUN-2000; 2000US-209467P.
PR 28-JUN-2000; 2000US-214886P.
PR 30-JUN-2000; 2000US-215135P.
PR 07-JUL-2000; 2000US-216647P.
PR 07-JUL-2000; 2000US-216880P.
PR 11-JUL-2000; 2000US-217487P.
PR 11-JUL-2000; 2000US-217496P.
PR 14-JUL-2000; 2000US-218290P.
PR 26-JUL-2000; 2000US-220963P.
PR 14-AUG-2000; 2000US-224518P.
PR 14-AUG-2000; 2000US-224519P.
PR 14-AUG-2000; 2000US-225213P.
PR 14-AUG-2000; 2000US-225314P.
PR 14-AUG-2000; 2000US-225366P.
PR 14-AUG-2000; 2000US-225367P.
PR 14-AUG-2000; 2000US-225368P.
PR 14-AUG-2000; 2000US-225370P.
PR 14-AUG-2000; 2000US-225447P.
PR 14-AUG-2000; 2000US-225757P.
PR 14-AUG-2000; 2000US-225758P.
PR 14-AUG-2000; 2000US-225759P.
PR 18-AUG-2000; 2000US-226279P.
PR 22-AUG-2000; 2000US-226681P.
PR 22-AUG-2000; 2000US-226868P.
PR 23-AUG-2000; 2000US-227182P.
PR 23-AUG-2000; 2000US-227009P.
PR 30-AUG-2000; 2000US-228924P.
PR 01-SEP-2000; 2000US-229287P.
PR 01-SEP-2000; 2000US-229343P.
PR 01-SEP-2000; 2000US-229344P.
PR 01-SEP-2000; 2000US-229345P.
PR 05-SEP-2000; 2000US-229509P.
PR 05-SEP-2000; 2000US-229513P.
PR
PR 06-SEP-2000; 2000US-230437P.
PR 06-SEP-2000; 2000US-230438P.
PR 08-SEP-2000; 2000US-231242P.
PR 08-SEP-2000; 2000US-231243P.
PR 08-SEP-2000; 2000US-231244P.
PR 08-SEP-2000; 2000US-231413P.
PR 08-SEP-2000; 2000US-231414P.
PR 08-SEP-2000; 2000US-232080P.
PR 08-SEP-2000; 2000US-232081P.
PR 12-SEP-2000; 2000US-231968P.
PR 14-SEP-2000; 2000US-232397P.
PR 14-SEP-2000; 2000US-232398P.
PR 14-SEP-2000; 2000US-232399P.
PR 14-SEP-2000; 2000US-232400P.
PR 14-SEP-2000; 2000US-232401P.
PR 14-SEP-2000; 2000US-233063P.
PR 14-SEP-2000; 2000US-233064P.
PR 14-SEP-2000; 2000US-233065P.
PR 21-SEP-2000; 2000US-234223P.
PR 21-SEP-2000; 2000US-234274P.
PR 25-SEP-2000; 2000US-234997P.
PR 25-SEP-2000; 2000US-234998P.
PR 26-SEP-2000; 2000US-235484P.
PR 27-SEP-2000; 2000US-235834P.
PR 27-SEP-2000; 2000US-235836P.
PR 29-SEP-2000; 2000US-236327P.
PR 29-SEP-2000; 2000US-236367P.
PR 29-SEP-2000; 2000US-236368P.
PR 29-SEP-2000; 2000US-236369P.
PR 29-SEP-2000; 2000US-236370P.
PR 02-OCT-2000; 2000US-236802P.
PR 02-OCT-2000; 2000US-237037P.
PR 02-OCT-2000; 2000US-237038P.
PR 02-OCT-2000; 2000US-237039P.
PR 13-OCT-2000; 2000US-239935P.
PR 13-OCT-2000; 2000US-239937P.
PR 20-OCT-2000; 2000US-240960P.
PR 20-OCT-2000; 2000US-241221P.
PR 20-OCT-2000; 2000US-241785P.
PR 20-OCT-2000; 2000US-241786P.
PR 20-OCT-2000; 2000US-241787P.
PR 20-OCT-2000; 2000US-241808P.
PR 20-OCT-2000; 2000US-241809P.
PR 20-OCT-2000; 2000US-241826P.
PR 01-NOV-2000; 2000US-244617P.
PR 08-NOV-2000; 2000US-246474P.
PR 08-NOV-2000; 2000US-246475P.
PR 08-NOV-2000; 2000US-246476P.
PR 08-NOV-2000; 2000US-246477P.
PR 08-NOV-2000; 2000US-246478P.
PR 08-NOV-2000; 2000US-246523P.
PR 08-NOV-2000; 2000US-246524P.
PR 08-NOV-2000; 2000US-246525P.
PR 08-NOV-2000; 2000US-246526P.
PR 08-NOV-2000; 2000US-246527P.
PR 08-NOV-2000; 2000US-246528P.
PR 08-NOV-2000; 2000US-246532P.
PR 08-NOV-2000; 2000US-246609P.
PR 08-NOV-2000; 2000US-246610P.
PR 08-NOV-2000; 2000US-246611P.
PR 08-NOV-2000; 2000US-246613P.
PR 17-NOV-2000; 2000US-249207P.
PR 17-NOV-2000; 2000US-249208P.
PR 17-NOV-2000; 2000US-249209P.
PR 17-NOV-2000; 2000US-249210P.
PR 17-NOV-2000; 2000US-249211P.
PR 17-NOV-2000; 2000US-249212P.
PR 17-NOV-2000; 2000US-249213P.
PR 17-NOV-2000; 2000US-249214P.
PR 17-NOV-2000; 2000US-249215P.
PR 17-NOV-2000; 2000US-249216P.
PR 17-NOV-2000; 2000US-249217P.

PR 17-NOV-2000; 2000US-249218P.
PR 17-NOV-2000; 2000US-249244P.
PR 17-NOV-2000; 2000US-249245P.
PR 17-NOV-2000; 2000US-249264P.
PR 17-NOV-2000; 2000US-249265P.
PR 17-NOV-2000; 2000US-249297P.
PR 17-NOV-2000; 2000US-249299P.
PR 17-NOV-2000; 2000US-249300P.
PR 01-DEC-2000; 2000US-250160P.
PR 01-DEC-2000; 2000US-250391P.
PR 05-DEC-2000; 2000US-251030P.
PR 05-DEC-2000; 2000US-251988P.
PR 05-DEC-2000; 2000US-256719P.
PR 06-DEC-2000; 2000US-251479P.
PR 08-DEC-2000; 2000US-251856P.
PR 08-DEC-2000; 2000US-251868P.
PR 08-DEC-2000; 2000US-251869P.
PR 08-DEC-2000; 2000US-251989P.
PR 08-DEC-2000; 2000US-251990P.
PR 11-DEC-2000; 2000US-254097P.
PR 05-JAN-2001; 2001US-259678P.
XX (HUMA-) HUMAN GENOME SCI INC.
XX PA
XX Rosen CA, Barash SC, Ruben SM;
XX PI
XX WPI; 2001-476161/51.
DR N-PSDB; ABA06468.
XX
XX Isolated nucleic acid molecule encoding an inflammation-associated
PT polypeptide is used in preventing, treating or ameliorating a medical
PT condition -
XX
XX Claim 11; SEQ ID NO: 554; 859pp + Sequence Listing; English.
XX
XX The present invention provides human cDNAs, proteins and related genomic
CC DNAs. These can be used in the treatment of neural, immune system,
CC muscular, reproductive, gastrointestinal, pulmonary, cardiovascular,
CC renal and proliferative disorders and inflammation. The present sequence
CC is a protein of the invention.
XX
XX Sequence 1766 AA;
SQ

Query Match 98.7%; Score 9464.5; DB 22; Length 1766;
Best Local Similarity 99.0%; Pred. No. 0;
Matches 174; Conservative 0; Mismatches 1; Indels 17; Gaps 1;

QY 1 MECCRRATPTLLLLFLAFLLSSRTARSSEDRDGLWDANGPWSSECSRTCGGASYSIRRC 60
DB 22 meccrratptllllflaflllssrtarseedrdglwdawgpwssecsrtcggasysirrc 81

QY 61 LSKSCGGRNIRYRTCSNVDCPPPEAGDFRAQCSAHNDYKHHGQFYEWEFLVSNPDNPCCS 120
DB 82 lsksccegrniryrctsnvdcppeagdfraqqcsahndvkhggfyewflpsvndpnpccs 141

QY 121 LKCOAGTTLVVELAPKVLDTGTCYTESLDMCISGLCQIVGDHQLGSTVKEDNCVCNG 180
DB 142 lkcoagktlivelapkvldgtcritesldmcisglcqvigcdhqlgstvkedncvcng 201

QY 181 DGSCTRLVRGOYKSQLSATKSDDTVVAIPYGRHRIRLVKGPDPHLYLETKTQGTKGENS 240
DB 202 dgscrlvrqyqkslatskdsdtvvaipygrhrirlvkpgdpdhlylektlqgtkgens 261

QY 241 LSTGTFFLVNDSVDFQKPPDEILRMAGPLADFTVIRKIRNSGSDSTVOFIYQPIIHR 300
DB 262 lstgtfflvnsvdfqkpfdeilrmagpladftvirkirnsgsdstvqfifyqpiih 321

QY 301 WRETDFEPCSATCGGYQLTSECYDLSNRVVAQYCHYYPENIKPKPKLOECNLDPCP 360
DB 322 wretdfpcsatcggyqltsecydlrsnrvvadqychyyphenikpkpkloecnldpcp 381

QY 361 ASDGYKQIMPYDLYHPLPWEATPWTACSSCGGGTQSRVASCVEEDIQCHVTISVEWKC 420

Db 382 a-----rweatpwtacsscggyiqsraavscveediqghvtsveewkc 424
QY 421 MYTPKMPIAQCNIFFDCPKWLAQEWSPCTVTCGGGLRVRVVLCDHGRMHTGGSPKTKP 480
Db 425 mytpkmpiadqcniffdcpkwlagewspctvtcggglrvrvvvlcdhrgmhtggspktpk 484
QY 481 HIKEECIVPTPCYKPKKLPVEAKLPWFKAQAELEGAASVEEPSFTPEAWSACTVTCGV 540
Db 485 hikeecivptpcykpklpveaklpwfkaqeleeaavseepfipxawsactvtcgv 544
QY 541 GTQVRIVRQCQVLLSFSQSADLPIDECEGPKPASORACYAGPCSGEIPFENPDTDFLFG 600
Db 545 gtqvrivrqcqvllsfsqsadlpidecegpkpasqracysagpcsgelpefnpdetdglf 604
QY 601 GLODFEYDWEYEGFTKCSFSCGGGVQEAUVSLNKTQREPAENLCVTSRRPQLLKS 660
Db 605 glodfelydweyegftkcsescggvgqeaavscinkqtrepaeenlcvtstrppqlks 664
QY 661 CNLDPCPARWEIGKWSPCSLTCGVGLQTRDVFCSHLLSREMNETVILADELCRQPKPSTV 720
Db 665 cnldpcparweigkwspscltcgvglqtrdvfcshllsremnetviladelcrqpkpstv 724
QY 721 QACNRFNCPNWPYPAQWQPCSRCTCGGQVQKREVLCKORMADGSELELPETFCASASKPACQ 780
Db 725 qacnrfncppawypaqwpcsrctcgvgvqkrevlckqrmadgsflelpetfcasaskpacq 784
QY 781 QACKKDDCPSEWLLSDWTECSTCGEGTQTRSATCRKMLKTGLSTVNSVNSLCPPLPSSS 840
Db 785 qackkddcpsewllsdwrecstcgegtqtrsalkmlktglstvnstlcpllpfss 844
QY 841 IRPCLMATCARPGRPSTKHSPIHAAARKVYIOTRRQRKLFHVGGFAYLLPKTAVLURCP 900
Db 845 irpclmatcarprpstkhspihaaarkvyiotrrqrklhfvvggfayllpktaavlurcp 904
QY 901 ARVRKPLITWEGDQHLISSTVTVVAPFGYLYKTHRLKPSDAGVYTCGAPAREHFVIXL 960
Db 905 arvrkplitwkdghllsstvttvavfgylykthrlkpsdagvytcsaparehfvikl 964
QY 961 IGGNRKLVARPLSPRESEEVLAGRKGPKKALOTHHQNGIFSNKSKAERGLAANPGSR 1020
Db 965 iggnrklvarplsprseeevlagrkpgpkalqthkhqngifsnsgkaergrlaanpgsr 1024
QY 1021 YDVLVRLLEQGGWPCGELLASWEAQDSAEARNTTSEEDPGAEQVLLHLPFTMTVEQRLLD 1080
Db 1025 ydvlvrlleqggwpgelllasweaqdsaearttseedpgaeqvllhlpftmvtveqrlldd 1084
QY 1081 ILGNLSQOPEELRDLYSKHLVAQIAQBIFRSHLEHQDTLLKPSERRTSPVTLSPHKHVS 1140
Db 1085 ilgnlsqgpeelrdlyskhlvaqlaqeiffrshlehdqdtllkpserrtspvtlspkhvsg 1144
QY 1141 FSSSLRTSTGDAAGGSRRRPHRPTILRKISAAQOLISAEVWTHLGGTVALASGTLVLL 1200
Db 1145 fssslrtstgdaggsrrprhptilrkisaaqqlisasevvtlhlgctvalasgtlsvll 1204
QY 1201 HCEAIGHPRPTISWANGEEVQSDRILLQPDLSLOLAPVEADVGFTYCNATNALGYDS 1260
Db 1205 hceai ghprptiswangeevqsdrlillqpdslqlapveadvgyftcnaatnalgyds 1264
QY 1261 VSTAVTLAGKPLVKTSMVTINTEKPAVTVDIGSTIKTVQGVNVTINCOVAGVPEAEVTW 1320
Db 1265 vstiavtlagkplvktsmvtintekpavtvdigstiktvgvvnvtincqvagvpeaevtw 1324
QY 1321 FRNKSILGSPHHLHEGSLLLTINVSDDOGLYSCRAANLHGLTESTQLLLDPPQVPTQL 1380
Db 1325 frnksilgsphhhegslldtinvssddqglyscraanlhgeltstqlldppqvptql 1384
QY 1381 EDIRALLAATGPNLPSVLTSPQTQLVDPGNSALLGCPIKCHPVPNTWFGHGGQPIVTA 1440
Db 1385 edirallaatgpnlpvltspqltqlvdpngsallgcpikgbpvntwfhggqpivta 1444
QY 1441 TGLTTHILAAQGLLOVANLSSGGSGEFCIAONEAGVLMOKASIVLQDWWSDRLATCS 1500

Db 1445 tglthhlaagqilqlvanlsgsgqfscslaqneagvlmqkaslviqdywsvdrlatcs 1504
QY 1501 ASCNGVQOQPRCLNSDEVNPAHCAGKVRPAVOPICNRRDCPSRMVMTWSACTRS 1560
Db 1505 ascngvqvqprcllnstevnpahcagkvrpavpapiacnrrdcpsrmvmtwsactrs 1564
QY 1561 CGGGVTRRTVCOKLKASGISTPVSNDMCTQVAKRPVDTQACNQQLCWEAFSSWGQCNG 1620
Db 1565 cgggvtrrvvtcqklkaglistpvsndmctqvakrpvdtqacnqqlcweafsswgqng 1624
QY 1621 PCIGPHLAVQHRVFCQTRDGIITLPEQCSALPRPVSTQNCWSEACSVHVRVSLWLICTA 1680
Db 1625 pcigphlavqhrvfqctrdgitlpeqcsalprpvstqncwseacsvhvrslwlicta 1684
QY 1681 TCQNYGFSQSRVSCVHARTNKAVPEHLCSMGPRPANQRCNITPCENMECRDTRYCEKV 1740
Db 1685 tcqnygfsrrvecvhartnkavpehlcsmgprpanqrcnitpcenmecrdttrycekv 1744
QY 1741 KQKICQLSOFKSRCCGTCGKA 1762
Db 1745 kgiklqlsgfksrccgtcgka 1766

RESULT 3

AAB01431
ID AAB01431 standard; Protein: 874 AA.

XX AC AAB01431;

DT 20-OCT-2000 (first entry)

XX DE Human TANGO 224 (form 2).

XX TANGO; 128; 140; 197; 212; 213; 224; 239; modulating agent; asthma;
KW graft versus-host diseases; rheumatoid arthritis; psoriasis;
KW inflammatory bowel disease; septic shock; ulcerative colitis;
KW Crohn's disease; chronic myelogenous leukemia; cancer; liver
KW disease; Hodgkin's disease; osteoarthritis; Lyme's disease;
KW cachexia; autoimmune disease; myasthenia gravis; autoimmune diabetes;
KW systemic lupus erythematosus; transgenic animal; diagnosis;
KW prognosis; prophylactic; therapeutic; human.

XX OS Homo sapiens.

XX WN 2000039284-A1.

XX PD 06-JUL-2000.

XX PF 23-DEC-1999; 99WO-US31025.

XX PR 30-DEC-1998; 98US-0223546.

XX PS (MILL-) MILLENNIUM PHARM INC.

XX PA Holtzman DA;

XX PI WPI; 2000-465743/40.

XX DR N-PSDB; AAA47482.

XX Novel nucleic acid sequences encoding TANGO-128, 140, 197, 212, 213,
PT 224 and 239 polypeptides useful for the treatment of asthma, rheumatoid
PT arthritis, psoriasis and autoimmune diseases

XX Claim 8; Fig 30; 209pp; English.

XX Nucleic acids encoding TANGO polypeptides are useful as modulating
CC agents for regulating cellular processes like asthma, graft
CC versus-host diseases, rheumatoid arthritis, psoriasis, inflammatory
CC bowel disease, septic shock, ulcerative colitis, Crohn's disease,
CC chronic myelogenous leukemia, cancer, liver disease, Hodgkin's
CC disease, osteoarthritis, Lyme's disease, cachexia and autoimmune
CC diseases e.g. myasthenia gravis, autoimmune diabetes and systemic
CC lupus erythematosus. The nucleic acids are also useful for producing

CC transgenic animals and the TANGO polypeptides themselves. Partial
TANGO-128, 140, 197, 212, 213, 224, 239 sequences are useful in
CC forensic biology, for diagnostic assays, prognostic assays,
CC pharmacogenomics and for monitoring clinical trials. TANGO
CC polypeptides are suitable for both prophylactic and therapeutic
CC methods for treating a subject at risk of a disorder or having a
CC disorder associated with aberrant TANGO expression. A wide range
CC of cellular disorders can be treated.
XX
SQ Sequence 874 AA;

Query Match 50.0%; Score 4797.5; DB 21; Length 874;
Best Local Similarity 98.9%; Pred. No. 2.3e-315;
Matches 864; Conservative 1; Mismatches 0; Indels 9; Gaps 1;

QY 1 MECRRATPTLLFLAFLLLSSWTARSEDRDGLDAMGWPWSECSTCGGASYSLRRC 60
Db 1 mecrratptglllflaflillssrtarseedrdglwdawgpwsecstcggasyslrrc 60
QY 61 LSSKSCGRNLRVTCSENVDPPPEAGDFRAQOCSAHNDVKHGOFYEWLPVSNDDPNPCS 120
Db 61 lsskscegrnirvrtcsenvdpcpeagdfraqcsahndvkhgqfyewlpvsnddpnpcs 120
QY 121 LKQAKGTTLLVVELAPKVLDTGTRCYTESLDMCISGLCQIVGCDHQLGSTVKEDNCGVNG 180
Db 121 lkqakgttllvvelapkvldgttrcytesldmcisglcqiivgcdhqlgstvkedncgvng 180
QY 181 DGSTCRIVRGQYKSQLSATKSDDTVVAIPYGSRHILVLKGPDLHYLETTLQSTKGENS 240
Db 181 dgstcrivrvgqyksqlsatksdtdtvaipygsrhilvlkghlylettlqstkgens 240
QY 241 LSSTGTFLVNSVDFDKFDPKKEILRMAGPLTADFIVKIRNSGASDSTVQFIFQPIIHR 300
Db 241 lsstgtflvnsdvdfkdpkkelirmagpltdafivkirnsgsadstvfifqpiih 300
QY 301 WRETDFPCSATCGGGYQLTSAECYDLRSNRVADQYCHYYPENIKPKLQECNLDPCP 360
Db 301 wretdfpcsatcggyqltsaecydlrsnrvadqychyyypenikpkqlqecnldpcp 360
QY 361 A-----SDGYKQIMPYDLYHPLPRWEATPWATGSSCGGGIQSRVSCVEEDTOGH 411
Db 361 aqrvgiivpnsgdykqimpydlyhplprweatpwatgsscgggigravscveediogh 420
QY 412 VTSVEWKMCTPKMPTAOPCNITFDCPKWLAQEWSPCTVTCGGRLRYRVILCDHRGHT 471
Db 421 vtsveewkcmtytpkmpiaqpcnifdcpkwlaqewspctvtcgggryrvilcdhrgmht 480
QY 472 GGCSPKTKPHKEECIVPTFCYKPKELPVEAKLPWFKAQLEEGAAVSEEPSFIFEAW 531
Db 481 ggcspktkphikeecivptfcykpkelpveaklpwfkaqleegaaavseepsfipeaw 540
QY 532 SACTVTGCVGTQVIRVRCQVLLSFSQSVDLPIDBCGPKPASORACYAGPCSEIPEFN 591
Db 541 sactvtcgvgtqvrvrcqvllsfsqsvadlpidecgpkpasacracypagpcseipefn 600
QY 592 PDETDGLFGLQDFDELYDWEYEGFTKCSGCGGVQAEVAVVCLNKOTREPAPENLCVTS 651
Db 601 pdetdglfglqdfdelydweyegftkcsescgvgvqaeavvclnkcrtrepapenlcvt 660
QY 652 RRPQLKSCNLDPCPARWEIGKWSPCSLTCGVGLQTRDYFCSHLLSRENETVILADEL 711
Db 661 rrpqllkscnldpcparweigkwspscltcgvglqtrdyfcsllsremnetviladel 720
QY 712 CRQPKPTVQACRNFPCPPAWYPAQWQPCSTCGGGVQKREVLCKQRMAGSGFLELPETF 771
Db 721 crqpkptvqacnrfncppawypaqwpcstcggvgvqkrevlckqrmagsgflepETF 780
QY 772 CSASKPACQACQKDDCPSFWLLSDWTFCSTCGEGTGTBSAICRKMILKTGLSVNVSTL 831
Db 781 csaskpacqacqkddcpsfwllsdwtfcstcsgegtgtbsaicrkmilktglstvnstl 840
QY 832 CPPLPFSSSIRPCMLATCARPGRPSTKHSPHIAA 865

QY 1311 AGVPAEAVTFWFKSKL-GSPHLLHEGSLLTNVSSDOGLYSCRAANLHGELTESTQLL 1369
Db 1324 kypgpntwtkrgslgnvslfngslilqnvslenegtyvciatnalgkavatsvln 1383
QY 1370 IIDPPQVPTQLEDIRA----LLAATGPNLPSVLTSPLGTQLVLDPNSALLGCPKIGHPV 1425
Db 1384 llerwpesrivilgghkkyilqatnrtns-----ndp-----tgepp 1422
QY 1426 PNITWPHGGQPIVTATGLTHHLLAAGQIIQVANLSGSGSQGEPSCLAQNEAGVLMQKASLV 1485
Db 1423 p----- 1423
QY 1486 IQDYWMSVDELATCSASCGNRGVQOPRLRCLL-NSTEVPNPAHCAGKVRPAVQPIACNRRD 1544
Db 1424 -qepfwepgnwshscatcghlariqpqcmangevsealcldhqlkplagfepcnird 1482
QY 1545 CFSRWMTWSWACTRSCGGVGTTRVTCOKLKASGISTPVSNDMCTQVAKRPVDTQACNQ 1604
Db 1483 cparftsvsqsvscsgegyhsrvtckrtkangltvqvvspracap-kdrplgrkpcfg 1541
QY 1605 QLCVEMAFSSWQCNGPCIGPHLAVQHOVFCQTRDITLPSQCSCALPRPVSTQNCWGE 1664
Db 1542 hpcvqw--epgnrcpgrcmgravrmmqgrhtacqhns----dsncddrkrptlrrnctsg 1595
QY 1665 ACSVHWRVSLWLTATCGNYGFSRRVCEVHARTNKAVPEHLCSWGPANWQRCNITP 1724
Db 1596 acdvcwhtpwkpctaacgr-gfgrskvdcihtrscpkvkrhcvqkxkpwiswhc-lgp 1653
QY 1725 CENMECDRTTRICEKVQKQLQLSOFKSRCCGTC 1759
Db 1654 scdrctdtdthycmfvkhlnlcsldrykdrccqsc 1688

RESULT 5

AAE10609
ID AAE10609 standard; Protein; 1691 AA.

XX AC AAE10609;

XX DT 10-DEC-2001 (first entry)

XX DE Human novel KIAA1233-like protein, NOV-2a.

XX KW Human; NOV-X protein; KIAA1233-like protein; STE20-like protein; tumour;
KW trypsin inhibitor-like protein; gene therapy; haematopoietic; illness;
KW immunological disorder; neurodegenerative disorder; Alzheimer's disease;
KW Parkinson's disease; immunomodulatory; pharmacogenomic; haemostatic;
KW human immunodeficiency virus; HIV; fertility disorder; neuroprotective;
KW cytotstatic; nootropic; anti-infertility; cancer; chromosome 15;
KW NOV-2a protein.

XX OS Homo sapiens.

XX FH Key Location/Qualifiers

FT Active-site 1..26

FT /label= Signal_peptide

FT Protein 27..1691

FT /note= "Human mature novel KIAA1233-like protein,
FT NOV-2a"

XX WO200162928-A2.

XX 30-AUG-2001.

XX 26-FEB-2001; 2001WO-US06151.

XX 25-FEB-2000; 2000US-0184951.

XX 28-FEB-2000; 2000US-0185548.

XX 01-MAR-2000; 2000US-0185967.

XX 18-APR-2000; 2000US-0197723.

XX 27-APR-2000; 2000US-0199957.

XX 23-FEB-2001; 2001US-0789390.

XX (CURA-) CURAGEN CORP.

XX Vernet CAM, Fernandes E, Shimmets RA, Macdougall J, Spaderna SK;

XX WPI; 2001-582051/65.

XX N-PSDB; AAD17757.

XX New isolated KIAA1233-like, STE20-like, or trypsin inhibitor-like
XX polypeptide for diagnosing and treating pathological disorders, such as
XX Parkinson's disease and for use in pharmacogenomics -

XX Claim 1; Page 22-23; 189pp; English.

XX The invention relates to novel human polypeptides referred as NOV-X
XX and their corresponding nucleic acid sequences. NOV-X collectively
XX include NOV-1, NOV-2a and NOV-2b which are novel KIAA1233-like
XX polypeptides, NOV-3a, NOV-3b, NOV-3c and NOV-3d which are novel
XX STE20-like polypeptides and NOV-4a, NOV-4b, NOV-4c, NOV-4d and NOV-4e
XX which are novel trypsin inhibitor-like polypeptides. NOV-X is used to
XX identify a potential therapeutic agent that can modulate its activity
XX and can be used for treating a pathology related to aberrant expression
XX or aberrant physiological interactions of NOV-X. NOV-X or its DNA is
XX used to determine the presence or predisposition to a disease associated
XX with altered levels of NOV-X. NOV-X, its DNA and its antibody are used
XX to treat or prevent a pathology associated with NOV-X. The pathological
XX states that can be treated or prevented are haematopoietic, cancer,
XX immunological, tumour, neurodegenerative (e.g. Alzheimer's and
XX Parkinson's disease), human immunodeficiency virus (HIV) illness and
XX fertility disorders. NOV-X and its DNA are used in pharmacogenomics for
XX predictive medicine. NOV-X DNA is used in gene therapy. The present
XX sequence is human novel KIAA1233-like protein, NOV-2a. NOV-2a gene is
XX located at chromosome 15.

XX Sequence 1691 AA;

Query Match 41.2%; Score 3953.5; DB 22; Length 1691;

Best Local Similarity 43.3%; Pred. No. 7.2e-258;

Matches 768; Conservative 265; Mismatches 555; Indels 187; Gaps 31;

QY 22 SSTARSEDRDGLMDAWGPWSECSRTCCGGASYSRLRCLSSKSCGRNIRYRTCSNVDC 81

Db 64 tsntsrtsdcdkgndawgdwscrtcggsasysrlrcrltgrnceggniryktsnhdh 123

QY 82 PPBAGDFRAOCCSAHNDYKHGOFYEWLPVSNPDNPDCSLKCOAKGTIVVELAPKVLG 141

Db 124 pdaeifraqcsayndvqyghyewlpryndpaapcalckhaqgnlvvelapkvldg 183

QY 142 TRCYTESLDMCISGLCQIVGCDHQLASTVAKEDNCGVCGNGDSTCRLVRGQYKSQLSATKS 201

Db 184 trcntdsldmcisgicqavgcdrglsgnakcdncgvcagdgstcrivrsgskshvapekr 243

QY 202 DDTWVAIPYSGRHIRLVKLPDHLIYETKTLQGTGKNSLSSTGTFLVDNSSVDFQKFPD 261

Db 244 eenviavlgrrsvritvkqphalfiesktlqgskgshsnspgvfvventtvefgqse 303

QY 262 KEILRMAGPLTADIVKIRNSGASDSTVQIFVQPIIHRWRRETDFFPCSATCGGQYLT 321

Db 304 rgtfkilpgplmadfifktrytaaksvvqffqyqishqwrqtdffpctvtcgggyqlns 363

QY 322 AECYDLRSNRVADQYCHYYPENIKPKQLQECNLDPCPASDGYKQIMPYDLVHPLPRWE 381

Db 364 aecvdrlkrvdpdhychypenivkpkpkkecsmdpcpsddgfkempdydhfqpiprwe 423

QY 382 ATPWTACSSSCGGGIQSRVAVSCVEEDIQGHVTSVEEWKCMYTPKMPIAQPCNIFDCPKWL 441

Db 424 hnpwtacsvscgggiqrsvfvcveesmhgeilqveewkcmypkpkvmqtcnlfidcpkwi 483

QY 442 AQPWSPCTVTCGGGLRYRVVVICIDHRGMHTGCGSPKPKPHKEECIVPTPCYKPKELPV 501

Db 484 amewsgctvtcgrglryrvvvlcinhrgehvggcnqplklhkeecvipcykpkpksvp 543

Db 32 nceqgnirkyktsnhdcpdaedfraagcsayndvqyghyewlpryndpaapcalkc 91
QY 124 QAKGTTLWELAPKVLDTGRTYTESLDMCISGLCQIVGCDHQLGASTVKEDNCGVNGDGS 183
Db 92 haqgnlvvelapkvldgtrcndtsldmcisglcqvagcdrlqgsnakedncgvcagdg 151
QY 184 TCRIVRGQYKOLSATKSDDTVAIPVGSRRHRLVLKGPDPHLYLTKTQTKGENSLSS 243
Db 152 tcrivrgqskshvpekreenevaplgsrsvritvkgpahlfiesktlqgskgehsfns 211
QY 244 TGFLVDNSVDFOKFDPKELTMAGPLTADFIKIRNSGSADSTVOFIFQPIIHWRE 303
Db 212 pgvfvcentvfeigsercfkfpplmadfifkrytaakdsdvffqypishqwrq 271
QY 304 TDFPSCATCGGGYQLTSAECYDLRSNRVADQYCHYYPENIKPKLQECNLDPCPASD 363
Db 272 tdfpctvcggyqlnsaecdirlrvpdyhchypenvkpkpkkecsmdpcpsd 331
QY 364 GYKQIMPYDLYHPLPRWEATPWTACSSCGGQISRAVSCVEEDIOGHVTSVEEWKCMYT 423
Db 332 gfkempdyhfpplprwehnpwtacsvscggqldrrsfvceesmhgeillqveewkcm 391
QY 424 PKMPIAOPCNIFDCPKWLADEWSPCTVTCGGLRYRVVLCIDHRGMUTGCGSPKTKPHIK 483
Db 392 pkpkmvmtcnlfcdpkwianewscvtcgrglrvvclnhrghenvgcnplkhlk 451
QY 484 EECVPPCKYKPEKLPVEAKLPKQAELEGAAYSEEFSTPEAWSACTVTCGVGTQ 543
Db 452 eecvlpkpkpkpsveaklpwkqaeleetriateptfipepsacsttcgpgvq 511
QY 544 VRIVRCOVILSFQSADLPIDECCKPKPASOR-----ACYAGPCSGEIPFENDETDGL 598
Db 512 vrevkcvllitqtetelpeececgkplpterpcilleadcsapasreldiplp----- 565
QY 599 FGLQDPDELYDEYEGFTKCESCGGVQEAUVVSCLNKOTREPAENLCTVSRPQLL 658
Db 566 -----edsettydwaygftpctatclghqcatavclhigtqvtndslcdmvrhpams 621
QY 659 KSNLDPCPARWIGKWSPLCTGVGLQTRDVFCSHLLSRENNETVILADELCRQPKPS 718
Db 622 qacntepcprrwhvsgwpcsatcgvgqlrdvyclh-----pgetpappee-crdekph 675
QY 719 TVOACNRFNCPWAYPAQWPCSTCGGQKREVLCQKRMADGSLFELPFCPSAKPA 778
Db 676 alqacnqfcdpwpwhleewgqctcgggtqnrvtcrqltdgdsfnlsdelcggpkas 735
QY 779 COQACKKDCPSEWLLSDTECSTCGEGTQTRSAICRKMKTGLSTVWNSLCPPLPFS 838
Db 736 shksartcdcpplhvgdwksvscvgvqgrkvqcqrlaakgrriplesmmcrdlpgf 795
QY 839 SSIRPCMATCAR-PGRPSTK---HSPHIAAARKVYIQTTRORKLHFVVGFAYLKPKA 894
Db 796 plvrscqmpcskiksemktklgeqgpqllsvrvylqtreetkrinltigsrayllpnts 855
QY 895 VVLCRCARRVKPLINWEKDGHLSLSTHTVAPFGVLKILHRLKPSDAGVYTSAGAPARE 954
Db 856 viikcpvrrfqlkqliwekgdrcqlqnskrigltksgsikhhgaapdigvyrclagsaqe 915
QY 955 HFVILKIGGNKRLVAPRSPREEVLAKRGKGPKEALQTHKHQNGIFNSGSAEKRGLA 1014
Db 916 tvvlkligtndrliarp-----alrepmreygmhdhseanslgtvwhk--mrqmw 963
QY 1015 ANFGSRY--DDLVSRL-----LEQGGWPGELLASWFAQDSAEENNTTSEEDPGAEQVLLH 1066
Db 964 nnkndlyldddhinsqpfllrallghcsnagsnswelknkqfeaaavkq--ga----- 1014
QY 1067 LPFTWTEQRRLDDILGNLSQOPE--ELRDLYSKHLVAQLAEIFRSHLEHQDTLLKPSE 1124
Db 1015 --ysmdtaq--fdeirnmqlmetgevsddlasqliyqlvaelakapqthmq--wrgiq 1068
QY 1125 RRTSPVTLSPHKHVSFGSSSLRTSSTGAGGSRRRPHRK-----PTILRKISAA 1173
Db 1069 eetpp-----aaqlr-----getsvsgqshaknsqkltfkpkqgvimrq-sqp 1111

QY 1174 QQLGASEVV--THLGOTVALASGTLISVLLHCBAIGHPRPTISWARGEEVQFSRLLQPD 1232
Db 1112 psifnktinsrignvtvitrtevinilcdlitpseacytwtkdgtllqpsvkiildgt 1171
QY 1233 DSLQILAPEADVGFTYTCNATNALGYDSVSIATVILAGKPLVKISMTVINTEKPAVTVDI 1292
Db 1172 gkigiqnptrkeqgiyecsvanhlgdsvesssvlyaeapvilsvernitkpehnhlsvv 1231
QY 1293 GSTIKTVQGVVNTINCQVAGVPEAEVWFRNKSCL-GSPHILHEGSLLLTNVSSSDQGLY 1351
Db 1232 ggiiveaalganvtircpvgkvpqpnitwkrsgslsgnvsllfngsllllgnvslenegy 1291
QY 1352 SCRAANLHGETESTQLLILDPQVPTOLEDIRA-----LLAATGPNLPSVLTSPLGTQLV 1407
Db 1292 vciatnalgkavatsvhlhlerwpesrivflgghkkyilqatntrtns----- 1340
QY 1408 LDPNSALLGCPKIKGHPVPNTWFGGQPIVTATGLTHHILAAGQILOVANLSGSGQGEF 1467
Db 1341 ndp-----Lgeppp----- 1349
QY 1468 SCLAQNEAGVLMQKASLVIOQYVWSVDRLATCSASCGNRGVQOQRLRCLL-NSTEVNPAH 1526
Db 1350 -----gepfwepgnwshcsatcghlgarigrpqcvmangqevseal 1390
QY 1527 CAGKVRPAVQPIACNRDCPSRWMTSWSACTRSCGGGVOTRRVTCOKLKASGISTPVSN 1586
Db 1391 cdhlgkplagfepcnidrcparwftswwscvsgcyehsrqvtckrtkangtvqvvspp 1450
QY 1587 DMCTQVAKRPVDTQACNOOLCVENAFSSWGQCNCGPIGPHLAVQHRQVFCOTRBDITLPS 1646
Db 1451 racap-kdrplgrkpcfgbpcvq--epgnrcpgrcmgravrmmqrhtacqhss---d 1503
QY 1647 EQCSALRPVSTQNCWSEACSVHVRVSLWTLCATCGNYGQSRVCEVHARTKNAPEH 1706
Db 1504 snccdrkrtlrrnctsgacdvchwtpgkpcctaagr-gfgrskvdcihtrscpkvkr 1562
QY 1707 LCSGMPANWQRCNITPCENNECRDITRYCEKVKQLKCOLSOFKSRCCGTC 1759
Db 1563 hcvqkkkpswrhc-lpscdrdetdtthycmfvkhlnlcsldrykqrccqsc 1614

RESULT 7
AAV87347
ID AAV87347 standard; Protein; 643 AA.
XX
AC AAV87347;
XX
DT 11-MAY-2000 (first entry)
XX
DE Human signal peptide containing protein HSP-124 SEQ ID NO:124.
KW Human; signal peptide-containing protein; HSP; diagnosis; cancer;
KW inflammation; cardiovascular disease; anticancer; anti-inflammatory;
KW antimicrobial; nootropic; neuroprotective; cardiovascular; hepatotropic;
KW antiasthmatic; gene therapy; cell proliferation; neurological disorder;
KW reproductive disorder; developmental disorder; arteriosclerosis;
KW cirrhosis; psoriasis; acquired immune deficiency syndrome; anaemia;
KW asthma; Crohn's disease; infection; Alzheimer's disease; schizophrenia;
KW Parkinson's disease; Huntington's disease; ovulatory defect;
KW muscular dystrophy.
XX
OS Homo sapiens.
XX
XX W0200000610-A2.
PN
XX 06-JAN-2000.
PD
XX 25-JUN-1999; 99WO-US14484.
PF
XX 26-JUN-1998; 98US-0090762.
PR 31-JUL-1998; 98US-0094983.
PR 01-OCT-1998; 98US-0102686.

PR 11-DEC-1998; 98US-0112129.
XX (INCY-) INCYTE PHARM INC.
XX
PI Lal P, Tang YT, Gorgone GA, Corley NC, Guegler KJ, Baughn MR;
PI Akerblom IE, Au-Young J, Yue H, Patterson C, Keddy R, Hillman JL;
PI Bandman O;
XX
XX WPI; 2000-160673/14.
DR N-PSDB; AA298232.
XX
XX New human signal peptide-containing proteins useful in treatment,
PT prevention and diagnosis of e.g. cancer, inflammation and
PT cardiovascular disease
XX
XX Claim 1; Page 243-244; 327pp; English.
XX
XX AA298109 to AA298242 encode AA298724 to AA298737 which represent the
CC human signal peptide-containing proteins HSPB-1 to HSPB-134. HSPBs have
CC anticancer, anti-inflammatory, antimicrobial, neurotropic, hepatotropic,
CC neuroprotective, cardiovascular and antiasthmatic activities, and can
CC be used in gene therapy. HSPBs can be used to treat or prevent disorders
CC associated with decreased activity or function of HSPB. Antagonists of
CC HSPB are used to treat or prevent disorders associated with increased
CC activity or function of HSPB. Such diseases include cell proliferation
CC (including cancer), inflammation, cardiovascular, neurological,
CC reproductive or developmental disorders, (e.g. arteriosclerosis,
CC cirrhosis, psoriasis, acquired immune deficiency syndrome, anaemia,
CC asthma, Crohn's disease, microbial or other infections, congestive or
CC ischaemic heart disease, Alzheimer's, Parkinson's or Huntington's
CC diseases, schizophrenia, ovulatory defects, muscular dystrophy). HSPB
CC nucleic acids can be used for the recombinant production of HSPB, for
CC detecting HSPB in standard hybridisation and amplification assays (for
CC diagnosis and monitoring), in gene therapy, as antisense,
CC triplex-forming or ribozyme therapeutics, for detecting related sequences
CC or genetic variations, and for chromosomal mapping. HSPB are also used to
CC raise specific antibodies (Ab) and to screen for agonists and
CC antagonists (potential therapeutic agents). Ab are used to diagnose, or
CC monitor, HSPB-related diseases (in usual immunoassays), as therapeutic
CC antagonists, in competitive drug screens, and for purification of HSPB
CC from natural sources.
XX
XX Sequence 643 AA;

Query Match 37.2%; Score 3565; DB 21; Length 643;
Best Local Similarity 100.0%; Pred. No. 3e-232;
Matches 642; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MCCRATPTGTLTLLFLAFLLSSRTARSEDRDGLDAPGWSCRTCGGASYSLRRC 60
Db 1 meccratptgtlltllflafllssrtarseedrdglwdawgpwsecrctcggasyslrcc 60
QY 61 LSSKSCGRNRYRTCSNVDPPAGDFRAQCSAHNDVHHGQFYEWLPVSNPDNPDCS 120
Db 61 lssksccgrnryrtcsnvdppagdfraqcsahndvhhghqfyewlpvsnpdnpdncps 120
QY 121 LKCAQKTTLLVELAPKVLDTGTRCYTESLDMCISGLCQIVGCDHQLGSTVKEDNCYVNG 180
Db 121 lkcaqkttllvvelapkvldtgtrcytesldmcisglcqivgchqlgstvkedncycvng 180
QY 181 DGSFCLRVGRQYSQLSATSDDTVAIPVGSRRHRLVLKCPDHLYLETKTLQGTKEGNS 240
Db 181 dgstcrlrvgrqysqlsatssddtvaipvgsrrhrlvlkcpdhllyletktlqgtkgens 240
QY 241 LSSSTGTFVLVNSVDFKFPDKETLRMAGPLTADFIKVRNSGADSTVQFIFVQPIIHR 300
Db 241 lssstgtfvlvnsvdfkfpdketlrmagpltdfivkrnsgadstvfifvqpiihrr 300
QY 301 WRETDFPCSATCGGQYLTSAECYDLRSNRVADQYCHYYPENIKPKPLQECNLDPCP 360
Db 301 wretdffpcsatcggytltsaecydlrsnrvadqychyyypenikpkplqecnldpcp 360

QY 361 ASDGYKOIMPYDLYHPLPRWEATPWTACSSSCGGGIGSRVSCVVEEDIQGHVTSVEWKC 420
Db 361 asdgykqimpydlyhplprweatpwtacssscgggigsravscvveedilqghvtsveewkc 420
QY 421 MYTPKMPDIAQPCNIFDCPKWLAQEWSPCTVTCGGLRVVVLCTIDHRGHTGGCSPKTKP 480
Db 421 mytpkmpdiapcnifdcpkwlagewspctvtcggglrvvvlctidhrghmtggcspktkp 480
QY 481 HIKEECIVPTPCYKPKKLPVEAKLPWFKAQEELEGAASVEEPSFIPEAWSACTVTTCGV 540
Db 481 hikeecivptpcykpkklpveaklpwfkaqeleegaavseepsfipeawsactvtcgv 540
QY 541 GTQVRIVRQCQVLLSFSQSVADLPIDCEGPKPASQACYAGPCSGEIPENPDFTDGLFG 600
Db 541 gtqvriivrqcqvllsfsqsvadlpidecegpkpasqacypasgcsgelpefnpdetdglfg 600
QY 601 GLQDFDELYDWEYEGFTKCSFSCGGGVQEAUVVSCLNKQTRP 642
Db 601 glqdfdeydweyegftkcsescgggvqeaavvsclnkqtrep 642
RESULT 8
ABBI0454
ID ABB10454 standard; Protein; 645 AA.
XX ABB10454;
AC ABB10454;
DT 10-JAN-2002 (first entry)
XX Human cDNA SEQ ID NO: 762.
DE Human; gene therapy; neural disorder; immune system disorder;
XX muscular disorder; reproductive disorder; gastrointestinal disorder;
KW pulmonary disorder; cardiovascular disorder; renal disorder;
KW proliferative disorder; inflammation.
XX Homo sapiens.
OS
XX WO200154474-A2.
PN
XX 02-AUG-2001.
PD
XX
PF 17-JAN-2001; 2001WO-US01349.
PR 31-JAN-2000; 2000US-179065P.
PR 04-FEB-2000; 2000US-180628P.
PR 24-FEB-2000; 2000US-184664P.
PR 02-MAR-2000; 2000US-186350P.
PR 16-MAR-2000; 2000US-189874P.
PR 17-MAR-2000; 2000US-190076P.
PR 18-APR-2000; 2000US-198123P.
PR 19-MAY-2000; 2000US-205515P.
PR 07-JUN-2000; 2000US-209467P.
PR 28-JUN-2000; 2000US-214886P.
PR 30-JUN-2000; 2000US-215135P.
PR 07-JUL-2000; 2000US-216647P.
PR 07-JUL-2000; 2000US-216880P.
PR 11-JUL-2000; 2000US-217487P.
PR 11-JUL-2000; 2000US-217496P.
PR 14-JUL-2000; 2000US-218290P.
PR 26-JUL-2000; 2000US-220963P.
PR 26-JUL-2000; 2000US-220964P.
PR 14-AUG-2000; 2000US-224518P.
PR 14-AUG-2000; 2000US-224519P.
PR 14-AUG-2000; 2000US-225213P.
PR 14-AUG-2000; 2000US-225214P.
PR 14-AUG-2000; 2000US-225266P.
PR 14-AUG-2000; 2000US-225267P.
PR 14-AUG-2000; 2000US-225268P.
PR 14-AUG-2000; 2000US-225270P.
PR 14-AUG-2000; 2000US-225447P.
PR 14-AUG-2000; 2000US-225757P.
PR 14-AUG-2000; 2000US-225758P.

PR 14-AUG-2000; 2000US-225759P.
PR 18-AUG-2000; 2000US-226279P.
PR 22-AUG-2000; 2000US-226681P.
PR 22-AUG-2000; 2000US-226868P.
PR 22-AUG-2000; 2000US-227182P.
PR 23-AUG-2000; 2000US-227009P.
PR 30-AUG-2000; 2000US-228924P.
PR 01-SEP-2000; 2000US-229287P.
PR 01-SEP-2000; 2000US-229343P.
PR 01-SEP-2000; 2000US-229344P.
PR 01-SEP-2000; 2000US-229345P.
PR 05-SEP-2000; 2000US-229509P.
PR 06-SEP-2000; 2000US-229513P.
PR 06-SEP-2000; 2000US-230437P.
PR 08-SEP-2000; 2000US-230438P.
PR 08-SEP-2000; 2000US-231242P.
PR 08-SEP-2000; 2000US-231243P.
PR 08-SEP-2000; 2000US-231244P.
PR 08-SEP-2000; 2000US-231413P.
PR 08-SEP-2000; 2000US-231414P.
PR 08-SEP-2000; 2000US-232080P.
PR 08-SEP-2000; 2000US-232081P.
PR 12-SEP-2000; 2000US-231968P.
PR 14-SEP-2000; 2000US-232397P.
PR 14-SEP-2000; 2000US-232398P.
PR 14-SEP-2000; 2000US-232399P.
PR 14-SEP-2000; 2000US-232400P.
PR 14-SEP-2000; 2000US-232401P.
PR 14-SEP-2000; 2000US-233063P.
PR 14-SEP-2000; 2000US-233064P.
PR 14-SEP-2000; 2000US-233065P.
PR 21-SEP-2000; 2000US-234223P.
PR 21-SEP-2000; 2000US-234224P.
PR 23-SEP-2000; 2000US-234997P.
PR 25-SEP-2000; 2000US-234998P.
PR 26-SEP-2000; 2000US-235484P.
PR 27-SEP-2000; 2000US-235834P.
PR 27-SEP-2000; 2000US-235836P.
PR 29-SEP-2000; 2000US-235836P.
PR 29-SEP-2000; 2000US-236327P.
PR 29-SEP-2000; 2000US-236367P.
PR 29-SEP-2000; 2000US-236368P.
PR 29-SEP-2000; 2000US-236369P.
PR 29-SEP-2000; 2000US-236370P.
PR 02-OCT-2000; 2000US-236802P.
PR 02-OCT-2000; 2000US-237037P.
PR 02-OCT-2000; 2000US-237038P.
PR 02-OCT-2000; 2000US-237039P.
PR 02-OCT-2000; 2000US-237040P.
PR 13-OCT-2000; 2000US-239935P.
PR 13-OCT-2000; 2000US-239937P.
PR 20-OCT-2000; 2000US-240960P.
PR 20-OCT-2000; 2000US-241221P.
PR 20-OCT-2000; 2000US-241785P.
PR 20-OCT-2000; 2000US-241786P.
PR 20-OCT-2000; 2000US-241787P.
PR 20-OCT-2000; 2000US-241808P.
PR 20-OCT-2000; 2000US-241809P.
PR 20-OCT-2000; 2000US-241826P.
PR 01-NOV-2000; 2000US-244617P.
PR 08-NOV-2000; 2000US-246474P.
PR 08-NOV-2000; 2000US-246475P.
PR 08-NOV-2000; 2000US-246476P.
PR 08-NOV-2000; 2000US-246477P.
PR 08-NOV-2000; 2000US-246478P.
PR 08-NOV-2000; 2000US-246523P.
PR 08-NOV-2000; 2000US-246524P.
PR 08-NOV-2000; 2000US-246525P.
PR 08-NOV-2000; 2000US-246526P.
PR 08-NOV-2000; 2000US-246527P.
PR 08-NOV-2000; 2000US-246528P.
PR 08-NOV-2000; 2000US-246532P.
PR 08-NOV-2000; 2000US-246609P.
PR 08-NOV-2000; 2000US-246610P.

PR 08-NOV-2000; 2000US-246611P.
PR 08-NOV-2000; 2000US-246613P.
PR 17-NOV-2000; 2000US-249207P.
PR 17-NOV-2000; 2000US-249208P.
PR 17-NOV-2000; 2000US-249209P.
PR 17-NOV-2000; 2000US-249210P.
PR 17-NOV-2000; 2000US-249211P.
PR 17-NOV-2000; 2000US-249212P.
PR 17-NOV-2000; 2000US-249213P.
PR 17-NOV-2000; 2000US-249214P.
PR 17-NOV-2000; 2000US-249215P.
PR 17-NOV-2000; 2000US-249216P.
PR 17-NOV-2000; 2000US-249217P.
PR 17-NOV-2000; 2000US-249218P.
PR 17-NOV-2000; 2000US-249244P.
PR 17-NOV-2000; 2000US-249245P.
PR 17-NOV-2000; 2000US-249264P.
PR 17-NOV-2000; 2000US-249265P.
PR 17-NOV-2000; 2000US-249297P.
PR 17-NOV-2000; 2000US-249299P.
PR 17-NOV-2000; 2000US-249300P.
PR 01-DEC-2000; 2000US-250160P.
PR 01-DEC-2000; 2000US-250391P.
PR 05-DEC-2000; 2000US-251030P.
PR 05-DEC-2000; 2000US-251988P.
PR 05-DEC-2000; 2000US-256719P.
PR 06-DEC-2000; 2000US-251479P.
PR 08-DEC-2000; 2000US-251858P.
PR 08-DEC-2000; 2000US-251868P.
PR 08-DEC-2000; 2000US-251869P.
PR 08-DEC-2000; 2000US-251989P.
PR 08-DEC-2000; 2000US-251990P.
PR 11-DEC-2000; 2000US-254097P.
PR 05-JAN-2001; 2001US-259678P.
XX
XX (HUMA-) HUMAN GENOME SCI INC.
XX
XX Rosen CA, Barash SC, Ruben SM;
PI
XX
XX
XX WPI; 2001-476161/51.
XX N-PSDB; ABA06676.
XX
XX Isolated nucleic acid molecule encoding an inflammation-associated
PT polypeptide is used in preventing, treating or ameliorating a medical
PT condition -
XX
XX Claim 11; SEQ ID NO: 762; 859pp + Sequence Listing; English.
PS
XX The present invention provides human cDNAs, proteins and related genomic
CC DNAs. These can be used in the treatment of neural, immune system,
CC muscular, reproductive, gastrointestinal, pulmonary, cardiovascular,
CC renal and proliferative disorders and inflammation. The present sequence
CC is a protein of the invention.
XX
SQ Sequence 645 AA;

Query Match 35.7%; Score 3423; DB 22; Length 645;
Best Local Similarity 99.8%; Pred. No. 1.2e-222;
Matches 636; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
QY 1126 RTSPVTLSPHKHVSFGFSSSLRTSTGDAGGSRPHRPHKPTILRKISAAQQLSASEVVTHL 1185
|||||
Db 9 rtsptlspkhkvsfgfssslrtstgdagggrrphrkptllrkisaaqqlsasevvthl 68
QY 1186 GQTVLASGTLSQLHCEAIGHPRPTISWARGSEVQFSDRIILQPDSDIQLILAPVEADV 1245
|||||
Db 69 gqtvalasgtlsqlhceaghprptiswargseevfssdrillqpddslqilapveadv 128
QY 1246 GFYTCNATNALGYDSVSIATVTLACKPLVKTSRMTVINTEKPAVTVDIGSTIKTVQGVNVT 1305
|||||
Db 129 gfytcnatanlgydsvsiavtvtagkplvktstrmtvintekpavtvdigstiktvggvnt 188

CC detect the presence of a tumour in a mammal by comparing the level of
CC expression of a PRO polypeptide in a test sample of cells from the animal
CC and a control sample of normal cells, whereby a higher level of
CC expression in the test sample indicates the presence of a tumour in the
CC mammal. Mammals include dogs, cats, cattle, horses, sheep, pigs, goats
CC and rabbits but are preferably human. The polypeptides can be used to
CC stimulate tumour necrosis factor (TNF) alpha release from human blood,
CC when contacted with it. A specific polypeptide can be used to stimulate
CC the proliferation or differentiation of chondrocyte cells. The PRO
CC proteins can be used to determine the presence of tumours and also
CC susceptibility to tumour development, particularly adrenal, lung, colon,
CC breast, prostate, rectal, cervical, or liver tumours, in mammalian
CC subjects. The oligonucleotide probes specific for the PRO nucleic acids
CC can be used for genetic analysis of individuals with genetic disorders.

XX Sequence 525 AA;

Query Match 30.4%; Score 2917; DB 22; Length 525;
Best Local Similarity 99.8%; Pred. No. 1.4e-188;
Matches 524; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 MECCRRATPGTLLFLAFLLSSRTARSEDRDGLWDAGPWSFCSRTCCGGASYSILRRC 60
DB 1 MECCRRATPGTLLFLAFLLSSRTARSEDRDGLWDAGPWSFCSRTCCGGASYSILRRC 60
QY 61 LSSKSCGGRNIRYTCNVDCPPEAGDPRACQCSAHNDVKHGFYEWLPVSNDDPNPCS 120
DB 61 LSSKSCGGRNIRYTCNVDCPPEAGDPRACQCSAHNDVKHGFYEWLPVSNDDPNPCS 120
QY 121 LKQAKGTTLVVELAPKVLDTGTCRYTESLDMCISGLCQIVGCDHQLGSTVKEDNCVCNG 180
DB 121 LKQAKGTTLVVELAPKVLDTGTCRYTESLDMCISGLCQIVGCDHQLGSTVKEDNCVCNG 180
QY 181 DSGTCLVRCQKSQLSATKSDTVAIPYGSRHRLVLKGPDLHLYLETKTQCKGENS 240
DB 181 DSGTCLVRCQKSQLSATKSDTVAIPYGSRHRLVLKGPDLHLYLETKTQCKGENS 240
QY 241 LSGTGFVLVDSNVDFOKPPDKILLRMAGPLTADFTVKIRNSGASDSTVOFIYQPIIHR 300
DB 241 LSGTGFVLVDSNVDFOKPPDKILLRMAGPLTADFTVKIRNSGASDSTVOFIYQPIIHR 300
QY 301 WRETDFPSCATCGGYYQLTSECYDLRNRVAVDGYCHYYPENIKPKKQLQECNLDPCP 360
DB 301 WRETDFPSCATCGGYYQLTSECYDLRNRVAVDGYCHYYPENIKPKKQLQECNLDPCP 360
QY 361 ASDGYKQIMPYDLYHPLPWEATPWTACSSCGGGIQRASVCEVEDIOGHVTSVEEWC 420
DB 361 ASDGYKQIMPYDLYHPLPWEATPWTACSSCGGGIQRASVCEVEDIOGHVTSVEEWC 420
QY 421 MYTPKMPIAQPCNIFDCPKWLAQWSPCTVTGCGGLRYRVVVICIDHRGHTGSCSKTKP 480
DB 421 MYTPKMPIAQPCNIFDCPKWLAQWSPCTVTGCGGLRYRVVVICIDHRGHTGSCSKTKP 480
QY 481 HIKEECIVTPTCYKPKKPLVPAKULPWFKQAQLEEGAAVSEEPS 525
DB 481 HIKEECIVTPTCYKPKKPLVPAKULPWFKQAQLEEGAAVSEEPS 525

RESULT 11
AAB44285
ID AAB44285 standard; Protein; 525 AA.

XX AAB44285;
XX AAB44285;
XX 08-FEB-2001 (first entry)
DT Human PRO1071 (UNQ528) protein sequence SEQ ID NO:301.
DE Human; secreted protein; transmembrane protein; PRO; EST; cytotstatic;
KW Human; expressed sequence tag; detection; cancer.
XX Homo sapiens.

XX WO200053756-A2.

PD 14-SEP-2000.

PF 18-FEB-2000; 2000WO-US04341.

XX 08-MAR-1999; 99WO-US05028.

PR 12-MAR-1999; 99US-0123957.

PR 29-MAR-1999; 99US-0126773.

PR 21-APR-1999; 99US-0130232.

PR 28-APR-1999; 99US-0131445.

PR 14-MAY-1999; 99US-0134287.

PR 23-JUN-1999; 99US-0141037.

PR 26-JUL-1999; 99US-0145698.

PR 29-OCT-1999; 99US-0162506.

PR 30-NOV-1999; 99WO-US28313.

PR 02-DEC-1999; 99WO-US28551.

PR 16-DEC-1999; 99WO-US28565.

PR 30-DEC-1999; 99WO-US30095.

PR 30-DEC-1999; 99WO-US31243.

PR 05-JAN-2000; 99WO-US31274.

PR 06-JAN-2000; 2000WO-US00219.

PR 06-JAN-2000; 2000WO-US00277.

XX 06-JAN-2000; 2000WO-US00376.

(GETH) GENENTECH INC.

XX Ashkenazi AJ, Baker KP, Botstein D, Desnoyers L, Eaton DL;

PI Ferrara N, Filvaroff E, Fong S, Gao W, Gerber H, Gerritsen ME;

PI Goddard A, Godowski PJ, Grimaldi CJ, Gurney AL, Hillan KJ;

PI Kljavin IJ, Kuo SS, Napier MA, Pan J, Paoni NF, Roy MA;

PI Shelton DL, Stewart TA, Tumas D, Williams PM, Wood WI;

XX WPI: 2000-611443/58.

DR N-PSDB; AAC78532.

XX Novel PRO polypeptides and polynucleotides used in detection methods,
PT to target bioactive molecules to specific cells, and to modulate
PT cellular activities -
XX Claim 12; Fig 120; 636pp; English.

XX AAC78458 to AAC78599 represent polynucleotide and EST (expressed
CC sequence tag) sequences which encode secreted or transmembrane PRO
CC polypeptides. The PRO polynucleotides and polypeptides have cytostatic
CC activity. The polynucleotides and polypeptides can be used for detecting
CC the presence of PRO polypeptides in samples, for linking bioactive
CC molecules to cells and for modulating biological activities of cells,
CC using the polypeptides for specific targeting. The polypeptide targeting
CC can be used to kill the target cells, e.g. for the treatment of cancers.
CC The polypeptide pairs provide specific targeting of bioactive molecules
CC to cells. AAC78600 to AAC78987 represent PCR primers and probes used in
CC the isolation of the PRO polynucleotide sequences.

XX Sequence 525 AA;

Query Match 30.3%; Score 2907; DB 21; Length 525;
Best Local Similarity 99.4%; Pred. No. 6.8e-188;
Matches 522; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

QY 1 MECCRRATPGTLLFLAFLLSSRTARSEDRDGLWDAGPWSFCSRTCCGGASYSILRRC 60
DB 1 MECCRRATPGTLLFLAFLLSSRTARSEDRDGLWDAGPWSFCSRTCCGGASYSILRRC 60

QY 61 LSSKSCGGRNIRYTCNVDCPPEAGDPRACQCSAHNDVKHGFYEWLPVSNDDPNPCS 120

DB 61 LSSKSCGGRNIRYTCNVDCPPEAGDPRACQCSAHNDVKHGFYEWLPVSNDDPNPCS 120

QY 121 LKQAKGTTLVVELAPKVLDTGTCRYTESLDMCISGLCQIVGCDHQLGSTVKEDNCVCNG 180

DB 121 LKQAKGTTLVVELAPKVLDTGTCRYTESLDMCISGLCQIVGCDHQLGSTVKEDNCVCNG 180

QY 181 DGSTCLVRGQYKQSLSATKSDTVAIPYGRHRLVLKGPDLHLYLETKTLQGTGKENS 240
Db 181 dgstclvrqyksqslsatksdttvvalpygrhrlvlkghpdhlyletktlqgtkgens 240
QY 241 LSSTGTLVDNSSVDFQKPPDKIELRMAGPLTADFTIVKIRNSGSADSTVQFTFYQPIIHR 300
Db 241 lsstgtflvndssvdfqkfpdkielrmagpltdftivkirnsgsadstvqfifyqpihr 300
QY 301 WRETDFPFCATCGGGYQLTSAECYDLRSNRVVADQYCHYYPENIKPKLOECNLDPCP 360
Db 301 wretdfpfcatscgggyqltsaecydlrsnrvvadqychyypenlkpkloecnldpcp 360
QY 361 ASDGYKQIMPYDLYHPLPRWEATPWTACSSCGGGIQSRVSCVEEDIQGHVTSVEEWKC 420
Db 361 asdgykqimpydlyhplprweatpwtacsscggyqsravscveediqghvtsveewkc 420
QY 421 MYTPKMPIAQCNIIDCPKWLAEHSPCTVTCGGLRYRVVLCIDHRGMHTGCCSPKTPK 480
Db 421 mytpkmpiaqcniidcpkwlqewspctvtcgggryrvvlicidhrgmhtgcsptktp 480
QY 481 HIKEECIVPTPCYKPKELPVEAKLPWFKQAELEGAASVEEPS 525
Db 481 hkeeciivptpcykpkelpveaklpwfkaqealeegaavseeps 525
RESULT 12
AAB72290
ID AAB72290 standard; Protein; 525 AA.
AC AAB72290;
XX
DT 14-MAY-2001 (first entry)
DE Human ADAMTS-R1 amino acid sequence.
DE
KW ADAMTS-N; disintegrin; metalloprotease; thrombospondin type I motif;
KW tumour cachexia; inflammation; dermatosparaxis; EDS-VIIC; angiogenesis;
KW Ehlers-Danlos syndrome type VIIC; articular cartilage erosion; human;
KW metastasis; embryogenesis; egg implantation; ADAMTS-R1.
XX Homo sapiens.
XX
XX WO200111074-A2.
XX
XX 15-FEB-2001.
XX
XX 03-AUG-2000; 2000WO-US21223.
XX
XX 06-AUG-1999; 99US-0369364.
XX (CLEV-) CLEVELAND CLINIC FOUND.
XX (APTE/) APTE S S.
XX (HURS/) HURSKAINEN T L.
XX (HIRO/) HIROHATA S.
XX
XX Aptc SS, Hurskainen TL, Hirohata S;
XX
XX WPI; 2001-159978/16.
XX DR N-PSDB; AAF63447.
XX
XX Murine and human 'A Disintegrin-like And Metalloprotease domain with
XX Thrombospondin type I motifs' proteins and the nucleic acids encoding
XX them, useful for treating e.g. tumours, inflammation and arthritis -
XX Claim 1; Fig 11; 181pp; English.
XX
XX This invention relates to murine and human ADAMTS-N (A disintegrin-like
XX and metalloprotease domain with thrombospondin type I motifs) proteins,
XX designated ADAMTS-5, 6, 7, 8, 9, 10 and R1. Also included in the
XX invention are cDNA sequences encoding the proteins, and antibodies
XX specific for the proteins. The nucleic acid sequences and proteins may be
XX used in the prevention, diagnosis and treatment of diseases associated

CC with inappropriate ADAMTS-N expression. Disorders that may be treated
CC using the nucleic acids, proteins and antibodies include, for example
CC tumour cachexia, inflammation, dermatosparaxis in cattle or Ehlers-Danlos
CC syndrome type VIIC (EDS-VIIC) in humans, erosion of articular cartilage
CC in arthritic (both inflammatory and non-inflammatory) disease,
CC angiogenesis, tumour growth and metastases, and they may also be used for
CC controlling embryogenesis and implantation of fertilised eggs. The
XX present sequence represents human ADAMTS-R1.
SQ Sequence 525 AA;
Query Match 30.3%; Score 2907; DB 22; Length 525;
Best Local Similarity 99.6%; Pred. No. 6.8e-188;
Matches 523; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
QY 1 MECCRRATPGTLLFLAFLLLSSRTARSEEDRDGLWDAMGPMSECSRTCGGGASYSLRRC 60
Db 1 meccrratpgtllflaflllsrtarseedrdglwdawgpmsecsrtcgggasnlrrc 60
QY 61 LSSKCEGRNIRYRTCSNVDCPEAGDFRAQCCSAHNDVKHHGQFYEWLPVSNDDPNPCS 120
Db 61 lsskscegrniryrctcsnvdcppeagdfraqccsahndvkhgqfyewlpvsndpnpes 120
QY 121 LKQAKGTTLVVELAPKVLDTGRTCYTESLDMGICISGLCQIVGCDHQLGCTVKEDNCGVCNG 180
Db 121 lkqakgttlvvelapkvldgtrcytesldmcisglcqi vgcdhqlgstvkedncvcng 180
QY 181 DGSTCLVRGQYKQSLSATKSDTVAIPYGRHRLVLKGPDLHLYLETKTLQGTGKENS 240
Db 181 dgstclvrqyksqslsatksdttvvaipygrhrlvlkghpdhlyletktlqgtkgens 240
QY 241 LSSTGTLVDNSSVDFQKPPDKIELRMAGPLTADFTIVKIRNSGSADSTVQFTFYQPIIHR 300
Db 241 lsstgtflvndssvdfqkfpdkielrmagpltdftivkirnsgsadstvqfifyqpihr 300
QY 301 WRETDFPFCATCGGGYQLTSAECYDLRSNRVVADQYCHYYPENIKPKLOECNLDPCP 360
Db 301 wretdfpfcatscgggyqltsaecydlrsnrvvadqychyypenlkpkloecnldpcp 360
QY 361 ASDGYKQIMPYDLYHPLPRWEATPWTACSSCGGGIQSRVSCVEEDIQGHVTSVEEWKC 420
Db 361 asdgykqimpydlyhplprweatpwtacsscggyqsravscveediqghvtsveewkc 420
QY 421 MYTPKMPIAQCNIIDCPKWLAEHSPCTVTCGGLRYRVVLCIDHRGMHTGCCSPKTPK 480
Db 421 mytpkmpiaqcniidcpkwlqewspctvtcgggryrvvlicidhrgmhtgcsptktp 480
QY 481 HIKEECIVPTPCYKPKELPVEAKLPWFKQAELEGAASVEEPS 525
Db 481 hkeeciivptpcykpkelpveaklpwfkaqealeegaavseeps 525
RESULT 13
AAE07868
ID AAE07868 standard; Protein; 845 AA.
XX
XX AAE07868;
XX
XX 01-NOV-2001 (first entry)
XX Novel human protein (NHP) #6.
XX Novel human protein; NHP; thrombospondin; gene therapy; cancer;
XX cytoskeletal; antisense therapy; angiogenesis; biological disorder.
XX Homo sapiens.
XX
XX WO200161011-A2.
XX
XX 23-AUG-2001.
XX
XX 15-FEB-2001; 2001WO-US05290.

XX 17-FEB-2000; 2000US-0183282.
XX (LEXI-) LEXICON GENETICS INC.
PI Donoho G, Scoville J, Turner CA, Friedrich G, Zambrowicz B;
PI Sands AT;
DR WPI; 2001-514776/56.
XX N-PSDB; AAD14369.
XX An isolated nucleic acid encoding a thrombospondin useful as a
PT hybridization probe and gene therapy treatments of cancer -
XX
PS Disclosure; Page 43-45; 56pp; English.
XX
CC The invention relates to novel human polynucleotides encoding proteins
CC that share sequence similarity with animal proteins having thrombospondin
CC repeats. NHPs are expressed in human cell lines, pituitary, lymph nodes,
CC prostate, testis, adrenal gland, uterus, foetal kidney, foetal lung and
CC gene trapped human cells. NHPs are thrombospondins useful for treating
CC biological disorders involving angiogenesis, cancer and development and
CC also in pharmacogenomic applications. NHPs are useful as a hybridisation
CC probe for screening libraries, assessing gene expression patterns and
CC also in gene therapy. Proteins having thrombospondin repeats act as
CC receptors, secreted extracellular matrix proteins and proteases. The
CC present sequence is novel human protein (NHP).
XX Sequence 845 AA;
XX
Query Match 29.3%; Score 2809.5; DB 22; Length 845;
Best Local Similarity 60.9%; Pred. No. 5.4e-181;
Matches 477; Conservative 115; Mismatches 170; Indels 21; Gaps 4;
XX 22 SRTARSEDRDGLDANGPWSECRSTCGGASYSIRRLCLSKSCGRNIRYETCSNVD 81
DB 64 tsrntrsdedkgnwdawgdwscdrctogggasysirrlcrtgrncoegnlrkytcsnhd 123
XX 82 PPEAGDFRAQCCSAHNDVKKHQFYEWLPVNDPNPCSLKQCAKGTTLVWEIAPKVL 141
DB 124 ppdaedfraqqcsayndvqgyhyewlpryndpaapcaalkchagqgnlvvelapkv 183
XX 142 TRCYTESLDMCISGLCQIVGCHQIGSTVKEKNCVCNGDSTCRIVRCQYKSLSATKS 201
DB 184 trcntsldmcisglcqvagcdrlqgsnakedncgvcagdstcrilvrggskshvpekr 243
XX 202 DDTVVAIPYGSRIHLVLKGDHVLKTLQGTGKNSISSTGTFLVDSNVDFOKFPD 261
DB 244 eenviavpgsvritvkgppahlfiesktlgsgkgehsfnpgvfvventtvefgrgse 303
XX 262 KEILRMAGPLTADFIKIRNSGSADSTVOFIFYQPIIHRWRETFPPCSATCGGGYQLTS 321
DB 304 rqtfkpglmadfikttrytaakdsvvqffypishqrtdffctvtcggvgqlns 363
XX 322 ACYDLRSNRVADQYCHYPENIKPKLQECNLDPASDGYKQIMPYDLYHPLPRWE 381
DB 364 acenvdirkrvvdphychypenvkpkllkecsmdpcpsdsgfkeimpdyhfpiprwe 423
XX 382 ATPWTACSSCGGGTQSRVAVSVEEDIOGHVTSVPEWKCMTPKMPTIAOPCNIFDCPKWL 441
DB 424 hnpwtacsvscgggqgrsvfcvcsmhgclqveckwmyapkpkmqtcnlfcdcpkw 483
XX 442 AQEWSPCTTCGGLRYRYVLICIDHRGMHTGGSPKTKPHIKECIVPTPCYKPKELPV 501
DB 484 amewsqctvtcgrlryrvvlcnhrgehvggcnplkhikeecvipipcykpkexspv 543
XX 502 EAKLPFKAOAELECAAYSEPSFTPEAWSACTVTCGVTQVRIYRCQVILLSFSQSVAD 561
DB 544 eaklpwkqaaleetriaeteeftipewasactctgpgvqvrevkcrvlltftqte 603
XX 562 LPIDCEGPKPASOR-----ACVAGPCSGEIPFNDETDGLFGGLQDFDELVDWEYEGF 616
XX
XX

DB 604 lpeeecepgkplptercilleadspasreldiplp-----edsettydweyagf 653
QY 617 TKCESCGGGVQEAIVVSCNLKQTRPAENLKVTSRRPQLLKSCNLDPARWEIGKWS 676
DB 654 tptctatclghqeaiavelhigtqvtvndslcdmvrppamsqacntepcprwhvgs 713
QY 677 PCSLTCGVLQTRDVPFCSHLSREMNETHVLADELCLROPKPSTVOACNRPNDPANYPAQ 736
DB 714 pcsatcgvgiqlrdvycllh-----pgetpappee-crdekphalqacnqldcpgpwhiee 767
QY 737 WQPCSTCGGCGVQKREVLCKQRMADGSELELPETFCASKPACQOACKKDDCPSEWLLSD 796
DB 768 wqpcstcggtgqrrrvtrcrltldgfslnlsdelcggpkasshskscartdcpphiavgd 827
QY 797 WTE 799
DB 828 wsk 830
XX
RESULT 14
AAE07869
ID AAE07869 standard; Protein; 771 AA.
XX
AC AAE07869;
XX
DT 01-NOV-2001 (first entry)
XX
DE Novel human protein (NHP) #7.
XX
KW Novel human protein; NHP; thrombospondin; gene therapy; cancer;
KW cytostatic; anticense therapy; angiogenesis; biological disorder.
XX
OS Homo sapiens.
XX
PN WO200161011-A2.
XX
PD 23-AUG-2001.
XX
PF 15-FEB-2001; 2001WO-US05290.
XX
PR 17-FEB-2000; 2000US-0183282.
XX
PA (LEXI-) LEXICON GENETICS INC.
XX
PI Donoho G, Scoville J, Turner CA, Friedrich G, Zambrowicz B;
PI Sands AT;
XX
DR WPI; 2001-514776/56.
DR N-PSDB; AAD14370.
XX
PT An isolated nucleic acid encoding a thrombospondin useful as a
PT hybridization probe and gene therapy treatments of cancer -
XX
PS Disclosure; Page 46-48; 56pp; English.
XX
CC The invention relates to novel human polynucleotides encoding proteins
CC that share sequence similarity with animal proteins having thrombospondin
CC repeats. NHPs are expressed in human cell lines, pituitary, lymph nodes,
CC prostate, testis, adrenal gland, uterus, foetal kidney, foetal lung and
CC gene trapped human cells. NHPs are thrombospondins useful for treating
CC biological disorders involving angiogenesis, cancer and development and
CC also in pharmacogenomic applications. NHPs are useful as a hybridisation
CC probe for screening libraries, assessing gene expression patterns and
CC also in gene therapy. Proteins having thrombospondin repeats act as
CC receptors, secreted extracellular matrix proteins and proteases. The
CC present sequence is novel human protein (NHP).
XX
SQ Sequence 771 AA;
XX
Query Match 27.3%; Score 2615.5; DB 22; Length 771;
Best Local Similarity 60.1%; Pred. No. 6.2e-168;
Matches 445; Conservative 110; Mismatches 165; Indels 21; Gaps 4;
XX

Db 361 tmvteqrrlddignlsqpeeelrldyskhlvaqlaqeifrshlehgdtllkpxsertsp 420
Qy 1130 vtlSPKHVSGFSSSLRTSSSTGDAGGSRPHRKPTILRKISAAOOLSASEVVTHLGQTV 1189
Db 421 vtlspkhvsgfssslrtsstgdaggsrrphrkptliirkisaagqlsasevvthlgqtv 480
Qy 1190 ALASGTLsv 1198
Db 481 alasgtlsv 489

Search completed: July 24, 2002, 04:17:42
Job time: 7030 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 23, 2002, 22:39:31 ; Search time 6233.88 seconds
(without alignments)
17754.665 Million cell updates/sec

Title: US-10-044-807-1
Perfect score: 5289
Sequence: 1 atggaatgctgcctcgggc.....gaacttggtgcaagcgtga 5289

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0
Searched: 1797656 seqs, 10463268293 residues
Total number of hits satisfying chosen parameters: 3595312

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl.*

- 1: gb.ba.*
- 2: gb.htg.*
- 3: gb.in.*
- 4: gb.om.*
- 5: gb.ov.*
- 6: gb.pat.*
- 7: gb.ph.*
- 8: gb.pl.*
- 9: gb.pr.*
- 10: gb.ro.*
- 11: gb.sts.*
- 12: gb.sy.*
- 13: gb.un.*
- 14: gb.vi.*
- 15: em.ba.*
- 16: em.fun.*
- 17: em.hum.*
- 18: em.in.*
- 19: em.mu.*
- 20: em.om.*
- 21: em.or.*
- 22: em.ov.*
- 23: em.pat.*
- 24: em.ph.*
- 25: em.pl.*
- 26: em.ro.*
- 27: em.sts.*
- 28: em.un.*
- 29: em.vi.*
- 30: em.htg_hum.*
- 31: em.htg_inv.*
- 32: em.htg_other.*
- 33: em.htgo_inv.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description

RESULT	1	AF176313	1571.4	29.7	1803	9	AF176313
LOCUS	AF176313	Homo sapiens	1369.2	25.9	1616	9	AF251058
DEFINITION	AF176313	Homo sapiens	1130	21.4	54193	9	AL591423
ACCESSION	AF176313	Homo sapiens	1130	21.4	205510	9	AC011701
VERSION	AF176313	Homo sapiens	1023.4	19.3	7260	6	AX235356
KEYWORDS	AF176313.1	GI:15099920	1020.2	19.3	5076	6	AX224809
SOURCE	human	human	956.2	18.1	4854	6	AX224823
ORGANISM	Homo sapiens	Homo sapiens	930.4	17.6	2538	6	AX224819
REFERENCE	1 (bases 1 to 1803)	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.	878.2	16.6	2523	9	AF237652
AUTHORS	Hirohata, S. and Apte, S.S.	Hirohata, S. and Apte, S.S.	866.4	16.4	2316	6	AX224821
TITLE	A novel member of ADAM-TS related gene, ADAM-TS1 (A Disintegrin-like And Metalloproteinase domain with Thrombospondin type I modules Related gene-1)	A novel member of ADAM-TS related gene, ADAM-TS1 (A Disintegrin-like And Metalloproteinase domain with Thrombospondin type I modules Related gene-1)	847.4	16.0	2175	6	AX224815
JOURNAL	Unpublished	Unpublished	835.4	15.8	8578	6	AX224825
REFERENCE	2 (bases 1 to 1803)	Hirohata, S. and Apte, S.S.	783.4	14.8	1953	6	AX224817
AUTHORS	Hirohata, S. and Apte, S.S.	Hirohata, S. and Apte, S.S.	676.8	12.8	2912	6	AX149471
TITLE	Direct Submission	Direct Submission	676.8	12.8	2912	6	AX149471
JOURNAL	Submitted (06-AUG-1999) Biomedical Engineering, The Cleveland Clinic Foundation, 9500 Euclid Avenue, ND20, Cleveland, OH 44195,	Submitted (06-AUG-1999) Biomedical Engineering, The Cleveland Clinic Foundation, 9500 Euclid Avenue, ND20, Cleveland, OH 44195,	554.8	10.5	1341	6	AX224811
REFERENCE	1 (bases 1 to 1803)	Hirohata, S. and Apte, S.S.	524.2	9.9	6303	6	AX235410
AUTHORS	Hirohata, S. and Apte, S.S.	Hirohata, S. and Apte, S.S.	522.6	9.9	6294	6	AX235354
TITLE	Direct Submission	Direct Submission	490.8	9.3	1119	6	AX224813
JOURNAL	Submitted (06-AUG-1999) Biomedical Engineering, The Cleveland Clinic Foundation, 9500 Euclid Avenue, ND20, Cleveland, OH 44195,	Submitted (06-AUG-1999) Biomedical Engineering, The Cleveland Clinic Foundation, 9500 Euclid Avenue, ND20, Cleveland, OH 44195,	357.2	6.8	68180	2	AC100465
REFERENCE	1 (bases 1 to 1803)	Hirohata, S. and Apte, S.S.	305.6	5.8	163163	9	AL353895
AUTHORS	Hirohata, S. and Apte, S.S.	Hirohata, S. and Apte, S.S.	238.6	4.5	213362	2	HS399M15
TITLE	Direct Submission	Direct Submission	225.2	4.3	78052	2	AC016062
JOURNAL	Submitted (06-AUG-1999) Biomedical Engineering, The Cleveland Clinic Foundation, 9500 Euclid Avenue, ND20, Cleveland, OH 44195,	Submitted (06-AUG-1999) Biomedical Engineering, The Cleveland Clinic Foundation, 9500 Euclid Avenue, ND20, Cleveland, OH 44195,	225.2	4.3	78052	2	AC016062
REFERENCE	1 (bases 1 to 1803)	Hirohata, S. and Apte, S.S.	218.2	4.1	78052	2	HS399M15
AUTHORS	Hirohata, S. and Apte, S.S.	Hirohata, S. and Apte, S.S.	213.6	4.0	213362	2	HS399M15
TITLE	Direct Submission	Direct Submission	207.4	3.9	5107	6	AX235413
JOURNAL	Submitted (06-AUG-1999) Biomedical Engineering, The Cleveland Clinic Foundation, 9500 Euclid Avenue, ND20, Cleveland, OH 44195,	Submitted (06-AUG-1999) Biomedical Engineering, The Cleveland Clinic Foundation, 9500 Euclid Avenue, ND20, Cleveland, OH 44195,	205.8	3.9	5107	6	AX235377
REFERENCE	1 (bases 1 to 1803)	Hirohata, S. and Apte, S.S.	205.8	3.9	5107	6	AX235379
AUTHORS	Hirohata, S. and Apte, S.S.	Hirohata, S. and Apte, S.S.	205.8	3.9	5107	6	AX235411
TITLE	Direct Submission	Direct Submission	204	3.9	68180	2	AC100465
JOURNAL	Submitted (06-AUG-1999) Biomedical Engineering, The Cleveland Clinic Foundation, 9500 Euclid Avenue, ND20, Cleveland, OH 44195,	Submitted (06-AUG-1999) Biomedical Engineering, The Cleveland Clinic Foundation, 9500 Euclid Avenue, ND20, Cleveland, OH 44195,	137.8	2.6	55276	2	AC110192
REFERENCE	1 (bases 1 to 1803)	Hirohata, S. and Apte, S.S.	131.6	2.5	324174	2	AC096334
AUTHORS	Hirohata, S. and Apte, S.S.	Hirohata, S. and Apte, S.S.	126.8	2.4	55276	2	AC110192
TITLE	Direct Submission	Direct Submission	121.8	2.3	15302	2	AC108657
JOURNAL	Submitted (06-AUG-1999) Biomedical Engineering, The Cleveland Clinic Foundation, 9500 Euclid Avenue, ND20, Cleveland, OH 44195,	Submitted (06-AUG-1999) Biomedical Engineering, The Cleveland Clinic Foundation, 9500 Euclid Avenue, ND20, Cleveland, OH 44195,	119.2	2.3	3604	3	AY061825
REFERENCE	1 (bases 1 to 1803)	Hirohata, S. and Apte, S.S.	118.6	2.2	16047	2	AC017996
AUTHORS	Hirohata, S. and Apte, S.S.	Hirohata, S. and Apte, S.S.	118.6	2.2	175179	3	AE009254
TITLE	Direct Submission	Direct Submission	118.6	2.2	311800	3	AE003781
JOURNAL	Submitted (06-AUG-1999) Biomedical Engineering, The Cleveland Clinic Foundation, 9500 Euclid Avenue, ND20, Cleveland, OH 44195,	Submitted (06-AUG-1999) Biomedical Engineering, The Cleveland Clinic Foundation, 9500 Euclid Avenue, ND20, Cleveland, OH 44195,	100	1.9	164879	9	AC022684
REFERENCE	1 (bases 1 to 1803)	Hirohata, S. and Apte, S.S.	100	1.9	181327	2	AC027807
AUTHORS	Hirohata, S. and Apte, S.S.	Hirohata, S. and Apte, S.S.	100	1.9	183976	2	AC073419
TITLE	Direct Submission	Direct Submission	97.8	1.8	2848	9	AF140674
JOURNAL	Submitted (06-AUG-1999) Biomedical Engineering, The Cleveland Clinic Foundation, 9500 Euclid Avenue, ND20, Cleveland, OH 44195,	Submitted (06-AUG-1999) Biomedical Engineering, The Cleveland Clinic Foundation, 9500 Euclid Avenue, ND20, Cleveland, OH 44195,	92	1.7	197748	2	HS570H19
REFERENCE	1 (bases 1 to 1803)	Hirohata, S. and Apte, S.S.	89.6	1.7	2445	6	AX327757

ALIGNMENTS

AF176313 1803 bp mRNA linear PRI 08-AUG-2001
Homo sapiens ADAM-TS related protein 1 (ADAMTSR1) mRNA, complete cds.

AF176313 1803 bp mRNA linear PRI 08-AUG-2001
Homo sapiens ADAM-TS related protein 1 (ADAMTSR1) mRNA, complete cds.

Phenology - N-term

```
FEATURES             Location/Qualifiers
     source            1..1803
                        /organism="Homo sapiens"
                        /db_xref="taxon:9606"
     gene              1..1803
                        /gene="ADAMTSR1"
                        /note="a disintegrin-like and metalloproteinase domain
                        with thrombospondin type I modules related gene-1"
     CDS                51..1628
                        /gene="ADAMTSR1"
                        /note="ADAM-TSRI; does not contain metalloproteinase
                        domain"
                        /codon_start=1
                        /product="ADAM-TS related protein 1"
                        /protein_id="AAK84170.1"
                        /db_xref="GI:15099921"
                        /translation="MECCRRATPGTLLFLAFLLSSRTARSEDRDLWDAMPWSE
                        CSRTCCGGGANSRLRCLSKSCGRNIRYTCNSVDCPEAGCFRAQQCSAHNDVXKH
                        GQYEWLPVSNDFDNCFLKQAKGTLLVVELAPKVDLGRCTESLSDMCISGICQIV
                        GDHQJGSLTKEDNCVNGDSTCLRVQYKQSLSKSDSDTVVAIPVSGSRHRLV
                        LKQPHLYLETTLQGTGKNSLSSTGLFLVNSYDFQKFPDKELRMAGPTADFI
                        VKIRNSGSADSTVQFYQPIIHRWRETDFPCCSATCGGQYQLTSAECYDLRNRVVA
                        DOYCHYYPINIKPKLOCNLDPCPASDYGKQIMPYDLYHPLPRWEATPWTACSSC
                        GGIOSRASVCEEDTQGHVTSVEWKCMYTPKMPAQPCNIPDCPKWLAQEWSPCTV
                        TGGQGLRYRVVICIDHRGMHTGGCSPKTHKEECIVPTPCYKPEKLPVEAKLPWF
                        KQAQLEEGEAAVSEEPS"
     variation          210..215
                        /gene="ADAMTSR1"
                        /note="changes translation from ala-asn to ser-tyr"
                        /replace="tcttaa"

BASE COUNT      475 a  460 c  457 g  411 t

Query Match      29.7%; Score 1571.4; DB 9; Length 1803;
Best Local Similarity 99.0%; Pred. No. 2e-271;
Matches 1581; Conservative 0; Mismatches 16; Indels 0; Gaps 0;

QY  1  atggaatgctgcggtcggaactctcggcacactgctctctcttcttgggtttctcgtc 60
DB  51  ATGGAATGCTGCCGTCGGGCAACCTCTGGCACACTGCTCTCTTCTGGGTTCTCTGCTC 110
QY  61  ctgagttccagaccgacgctcgaaggagaccggagccctatgggatgcctggggc 120
DB  111  CTGAGTTCCAGGACCGCACGCTCGAGGAGACCGGACGCCCTATGGGATGCTTGGGGC 170
QY  121  ccatggagtgaatgctcacgcacctgcgggggtggggcctcctactctctgagggcgtgc 180
DB  171  CCATGGAGTGAATGCTCACGCACCTGCGGGGTGGGGCGGCCCAACTCTCTGAGCGCTGC 230
QY  181  ctgagcagaagagctgtaggaagaaatatccgatcacagaacatcgagtaatgtggac 240
DB  231  CTGAGCAGCAAGAGCTGTGAAGGAAGAAATATCGATACAGAACATGCAGTAATGTGGAC 290
QY  241  tgccaccagaacagctgatttccagctcagcaatgctcagctcataatgatgtcaag 300
DB  291  TGCCACCAGAAGCAGGTGATTTCCGAGCTCAGCAATGCTCAGCTATATATGATGTCAAG 350
QY  301  caccatggcagtttatgaatggcttctcgtgtcctaatagccctgacaccccatgttca 360
DB  351  CACCATGGCCAGTTTATGAATGGCTTCTTCTGTCTAATGACCTGACACCCCATGTTC 410
QY  361  ctcaagtgcacaaagcaaacctcgtgtgttgaactagcacctaaaggtcttagat 420
DB  411  CTAAGTGCACAAAGCAAAACACCTCGTGTGTGTAATGATGATGATGATGATGATGAT 470
QY  421  ggtacgctgtctatcacagacttttgatattgtgcacatgagtggtttatgccaaattgt 480
DB  471  GGTACGCTGTCTATACAGATCTTTGGATATGTGCATCATGTTGTTATGCCAAATTTGT 530
QY  481  ggctgcgcatcacagctgggaagcacctgcaaggaagataactgtgggtctgcaacgga 540
|||||
```

```
DB  531  GGCCTGGCACTCACCAGCTGGGAAGCACCGTCAAGGAAGATAACTGTGGGGTCTGCAACGGA 590
QY  541  gatgggtccacctgcggtggttcgagggcagtgataataatcccgctccgcaacacaa 600
|||||
DB  591  GATGGGTCCACCTCGCGGTGTCGAGGGCAGTATAAATCCAGCTCTCCGCAACCAAAA 650
QY  601  tcgatatatactggttgtaattccctataggaagttagacatatcgcttctgttataaa 660
|||||
DB  651  TCGGATGATATCTGTGGTTCGAATTCCTATGGAAGTAGACATATTCGCTTGTCTTAAAA 710
QY  661  ggtcctgatacttatacttggaacacaaacccctccagggagactaaagtggaacagct 720
|||||
DB  711  GGTCCTGATCACTTATATCTGGAACCAAAACCCCTCCAGGGAGACTAAAGGTGAAACAGT 770
QY  721  ctcagctccacaggaactttccttctgtggacaattctagtgtggactccagaaattcca 780
|||||
DB  771  CTCAGCTCCACAGCAACTTTCCTTGTGGACAATCTAGTGTGGACTTCCAGAAATTTCCA 830
QY  781  gacaaagagatactgagaatggtctggaccactcagcagagatttcaattgtcaagattcgt 840
|||||
DB  831  GACAAAGAGATAC TGAGAATGGCTGGACACTCAACAGAGATTTCATTTGTCAAGATTCTGT 890
QY  841  aactcgggctcgcctgcagctacagtcacagtcacagttcatcttctatcaacccatcatccacga 900
|||||
DB  891  AACTTCGGGCTCCGCTGACAGTACAGTCCAGTTTCATCTTATCAACCCCATCATCCACCGA 950
QY  901  tggaggagagcggatttcttctctgtcagcaacctgtggagaggttatcagctgaca 960
|||||
DB  951  TGGAGGGAGACGGATTTCTTCTTGTCTCAGCAACCTGTGGAGGAGGTTATCAGCTGACA 1010
QY  961  tcggtcgaagtctacgatactctgagagaaacggtggtgtgtgtgacccaatactgtcaatat 1020
|||||
DB  1011  TCGGCTGAGTGTCTAGATCTGAGGAGCAACCGTGTGGTGTGTGCTGACCAATATCTGCTACAT 1070
QY  1021  taaccagagaacatacaaaccccaaaccaagcttcagagagtcgaacttgatctctgtcca 1080
|||||
DB  1071  TACCCAGAGAACATCAAAACCCAAAGCTTCAGGAGTCCAACTTGGATTCCTTTGTGTTCCA 1130
QY  1081  gccagtgacggatataaagcagatcatgctcttatgacctctaccatcccttctcctcgtgg 1140
|||||
DB  1131  GCCAGTGACGGATACAAAGCAGATCATGCTTATGACCTTACCATCCCTTCTCTCGGTGG 1190
QY  1141  gagccaccacatgagacgctgctcctcctcgtgtgggggggggcatccagagcgggca 1200
|||||
DB  1191  GAGGCCACCCCATGGACCGCTGCTCTCTCTGTTGGGGGGGGGCATCCAGAGCGGGGCA 1250
QY  1201  gtttctctgtgtggagagagacatccaggggcatgcaacttcagtggaagagtggaatgc 1260
|||||
DB  1251  GTTTCTCTGTGTGGAGGAGGACATCCAGGGGCATGTCACCTTCAGTGGAGAGTGGAAATGC 1310
QY  1261  atgtacacccctaagatgcccctgcgcagccctgcaacatcttttgactgccttaattgg 1320
|||||
DB  1311  ATGTACACCCCTAAGATGCCCATCGCGAGCCCTGCAACATTTTGTGCTGCTTAAATGG 1370
QY  1321  ctggcagagagtggtctcctgcacagtgacatgtggcagggcctcagatacctgtg 1380
|||||
DB  1371  CTGGCAGAGGAGTGTCTCTCTGTCAGTGACGTGTGGCGAGGGCCTCAGATACCGTGTG 1430
QY  1381  gtctctgtgcaLcgaccatLcgagggaatgcacacagagagctgtagccccaaaaaacaaagccc 1440
|||||
DB  1431  GTCCTCTGTCATCGACCATCGAGGAATGCACACAGAGAGGCTGTAGCCCCAAAAAACAAAGGCC 1490
QY  1441  cacataaaagaggaatgcatactgtaccactccctcgtctataaaccccaagagaactcca 1500
|||||
DB  1491  CACATAAAAGAGGAATGCATCGTACCCACTCCCTGCTATAAACCCAAAGAGAAACTTCCA 1550
QY  1501  gtcagggcccaagtggtccatggttcaacaagctcaagagctagaagagaggtgctgtg 1560
|||||
DB  1551  GTCAGGGCCAAAGTTGCCATGTTTCAACAAAGCTCAAGAGCTAGAGAGAGGAGCTGCTGTG 1610
QY  1561  tcagagagagcctcgttctaatcccgagagcgtgggtcgg 1597
|||||
DB  1611  TCAGAGGAGCCCTCGTAAGTTGTAAAGACACAGACTG 1647
```


LOCUS AC011701 205510 bp DNA linear PRI 27-FEB-2001
 DEFINITION Homo sapiens 12 BAC RP11-937121 (Roswell Park Cancer Institute
 Human BAC Library) complete sequence.
 AC011701
 VERSION AC011701.22 GI:13129402
 KEYWORDS HTG.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 1 (bases 1 to 205510)
 Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,
 Alsbrooks,S.L., Amarantunge,H.C., Are,J.R., Banks,T., Barbaria,J.,
 Benton,J., Bimaga,K., Blankenburg,K., Bonnin,D., Bouck,J.,
 Bowie,S., Brileva,M., Brown,B., Brown,M., Bryant,N.P., Buhay,C.,
 Burck,P., Burckett,C., Burrell,K.L., Byrd,N.C., Carron,T.F.,
 Carter,M., Cavazos,S.R., Chacko,J., Chavez,D., Chen,G., Chen,R.,
 Chen,Z., Chlidu,J., Chowdhry,I., Christopoulos,C., Cleveland,C.D.,
 Cox,C., Coyle,M.D., Dathorne,S.R., David,R., Davila,M.L., Davis,C.,
 Davy-Carroll,L., Dederich,D.A., Delaney,K.R., Delgado,O.,
 Denn,A.L., Ding,Y., Dinh,H.H., Douthwaite,K.J., Draper,H.,
 Dugan-Rocha,S., Durbin,K.J., Earnhart,C., Edgar,D., Edwards,C.C.,
 Elhaj,C., Emerling,S., Escotio,M., Falls,T., Ferraguto,D.,
 Flagg,N., Ford,J., Foster,P., Frantz,P., Gabisi,A., Gao,J.,
 Garcia,A., Garner,T., Garza,N., Gill,R., Gorrell,J.H., Guevara,W.,
 Gunaratne,P., Hale,S., Hamilton,K., Han,J., Harris,C., Harris,K.,
 Hart,M., Havlak,P., Hawes,A., Hernandez,J., Hernandez,O.,
 Hodgson,A., Hoques,M., Holloway,C., Hollins,B., Homsí,F.,
 Howard,S., Huber,J., Hulyk,S., Hume,J., Ioshikhes,I., Jackson,L.E.,
 Jacobson,B., Jia,Y., Johnson,R., Jolivet,S., Joudah,S.,
 Karissoun,E., Kelly,S., Khan,U., King,L., Korvah,J., Kovar,C.,
 Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lee,E., Lewis,L.C.,
 Lewis,L., Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W.,
 Lewis,H., Lozada,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R.,
 Ma,J., Maheshwari,M., Mapua,P., Marondel,I., Martin,R.,
 Martindale,A., Martinez,E., Massey,E., Mawhinney,E., McLeod,M.P.,
 Meador,M., Mei,G., Merscher,S., Metzger,M., Miller,A., Miner,G.,
 Miner,Z., Mitchell,T., Mohabbat,K., Montgomery,K.T., Morgan,M.,
 Morris,S., Moser,M., Neal,D., Nelson,D., Newton,J., Newton,N.,
 Nguyen,A., Nguyen,N., Nguyen,N., Nickerson,E., Nwokenkwo,S.,
 Ogih,M., Okwuonu,G., Oragunye,N., Oviedo,R., Pace,A., Payton,B.,
 Peery,J., Perez,L., Peters,L., Pickens,R., Primus,E., Pu,L.D.,
 Quiles,M., Ren,F., Rives,M., Rojas,A., Rojubokan,I., Rolfe,M.,
 Ruiz,S., Savery,G., Scherer,S., Scott,G., Shen,H., Shim,C.,
 Shoostari,N., Sisson,I., Sodergren,E., Sonaike,T., Sparks,A.,
 Stanley,H., Stone,H., Sutton,A., Svatek,A., Tabor,P., Tamerisa,A.,
 Tamerisa,K., Tang,H., Tansey,J., Taylor,C., Taylor,T., Telford,B.,
 Thomas,N., Thomas,S., Usmani,K., Vasquez,L., Vera,V., Villalon,D.,
 Vinson,R., Wall,R., Wang,S., Ward-Moore,S., Warren,R.,
 Washington,C., Watlington,S., Williams,G., Williamson,A.,
 Wlezyk,R., Wooden,S., Worley,K., Wu,C., Wu,Y., Wu,Y.F., Zhou,J.,
 Zorrilla,S., Kucherlapati,R. and Gibbs,R.
 Direct Submission
 2 (bases 1 to 205510)
 Worley,K.C.
 Direct Submission
 Submitted (12-OCT-1999) Human Genome Sequencing Center, Department
 of Molecular and Human Genetics, Baylor College of Medicine, One
 Baylor Plaza, Houston, TX 77030, USA
 3 (bases 1 to 205510)
 Worley,K.C.
 Direct Submission
 Submitted (27-FEB-2001) Human Genome Sequencing Center, Department
 of Molecular and Human Genetics, Baylor College of Medicine, One
 Baylor Plaza, Houston, TX 77030, USA
 On Feb 27, 2001 this sequence version replaced gi:13096014.
 INFORMATION: <http://www.hgsc.bcm.tmc.edu/> or email
gc-help@bcm.tmc.edu

CLONE LENGTH: This sequence does not necessarily represent the
 entire insert of this clone. Overlapping regions of clones are only
 sequenced and submitted once, so the sequence for the remainder of

the insert may be found in the record for the adjacent clones.
 Overlapping clones are noted at the beginning and end of the
 Features listing.

ANNOTATION OF FEATURES:

STSs are identified using ePCR (Genome Res. 7:541-550) searches
 of a local database that includes entries from dbSTS, GDB, and
 local mapping efforts.
 Repeats are identified using RepeatMasker (A. Smit and P. Green,
 unpublished.) for Human and Mouse sequences.

Genes and Region of sequence similarity are identified by BLAST
 (Nuc. Acids Res. 25:3389-3402) similarity (expect < 1e-34) to the
 EST and cDNA sequences. Genes demonstrate at least two exons
 flanked by consensus splice sites that maintained sequence
 continuity across the splice junctions. Sequences that are not
 identical matches are annotated as similar.

SEQUENCING READ COVERAGE: Sequencing is completed to a minimum
 standard of double strand coverage with a minimum of 2 clones and 2
 reads with no ambiguities or 2 chemistries with a minimum of 2
 clones and 3 reads with no ambiguities. If the sequence quality for
 a region does not meet this standard, it will be indicated in the
 annotation as Low Coverage.

QUALITY OF INDIVIDUAL BASES: This sequence meets stringent quality
 standards - estimated error rate less than 1 per 10,000 bases.
 Reports of lowest quality individual bases and measures of base
 quality are listed below. Description of the metrics can be found
 at URL: <http://gc.bcm.tmc.edu:8088/quality.info/genbank.annotation.html>.

QUALSTAT-REPORT-----

----- Summary Statistics -----
 Contig length: 205510
 Phrap values in estimate: 203901
 Average error rate (BCM-phrap estimate): 0.000164184
 Fraction of Phrap values less than 40 : 0.0264246
 Number of consensus changing edits: 29
 Number of N's in consensus : 0

----- Consensus changing edits -----

Position	Original*Context	Edited*Context
14478	taaaagccag(n)aaactatat	taaaagccag(g)aaactatat
14480	aaagccag(n)aaactatat	aaagccag(a)aaactatat
18344	tccttgaga(n)aaactatat	tccttgaga(g)aaactatat
18345	ctcttgaga(n)aaactatat	ctcttgaga(g)aaactatat
18346	ctcttgaga(n)aaactatat	ctcttgaga(g)aaactatat
30655	gcaagcaga(n)cagtcacaaat	gcaagcaga(a)cagtcacaaat
48571	taaaacactt(n)tgctgcccc	taaaacactt(c)tgctgcccc
85905	ctgggttg(g)ttttttttt	ctgggttg(g)ttttttttt
87055	aaacatgaat(n)aaacatata	aaacatgaat(c)aaacatata
96536	ccattttcc(n)ttttttttt	ccattttcc(t)ttttttttt
124767	ggcaataaca(n)nnactctgt	ggcaataaca(c)nnactctgt
124768	gcaataaca(n)nnactctgt	gcaataaca(c)nnactctgt
124769	gcaataaca(n)nnactctgt	gcaataaca(c)nnactctgt
124849	gcaataaca(n)nnactctgt	gcaataaca(c)nnactctgt
125277	ttatatgta(n)aaagagaag	ttatatgta(t)aaagagaag
125358	aatctcttg(g)gaggaacaa	aatctcttg(t)gaggaacaa
128124	ccagagaggt(n)tggttacga	ccagagaggt(c)tggttacga
128126	ccagagaggt(n)tggttacga	ccagagaggt(c)tggttacga
128206	tgagggtaca(n)agctagtgg	tgagggtaca(g)agctagtgg
136767	gctcaagtt(n)caggtatgaa	gctcaagtt(c)caggtatgaa
153400	tcagctctcc(n)tcacttagc	tcagctctcc(c)tcacttagc
159220	cccaagacct(n)ttgaagatca	cccaagacct(c)ttgaagatca
167658	atttatttt(n)nnctacatg	atttatttt(g)nnctacatg
167659	tttatttt(n)nnctacatg	tttatttt(g)nnctacatg
186539	agctggagga(n)ggaataactc	agctggagga(g)ggaataactc
186541	ctggggaggg(n)aaaaactcta	ctggggaggg(g)aaaaactcta
196467	ttaaaattac(t)tgtaaaatct	ttaaaattac(t)tgtaaaatct
196468	taaaattac(n)ntaaaatcta	taaaattac(t)ntaaaatcta

KEYWORDS

SOURCE human.
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
AUTHORS 1 (bases 1 to 7260)
Vernet,C.A., Fernandes,E., Shinkets,R.A., Macdougall,J. and Spaderna,S.K.
TITLE Polypeptides and nucleic acids encoding same
JOURNAL Patent: WO 0162928-A 3 30-AUG-2001;
Curagen Corporation (US)
FEATURES
source Location/Oualifiers
1..7260
/organism="Homo sapiens"
/db_xref="taxon:9606"
136..5211
/note="unnamed protein product"
/codon_start=1
/protein_id="CAC69812.1"
/db_xref="GI:15593890"
/translation="MASWTSPWMLIGVMHSPLOTTAETKSPGAYFLPERALSPQG
SLEDOTIGQFLTYRYDDQTSRNRSDKDNWDAGWSDSRTGGGASYSLRRC
LGRNCEGNIRYKTKSNIDCPDAEDFRAQCSAYNDVQYQHYEYEWLPRYNDPAP
CALKHAQGNLWELAPKLVSDTRCNDSLDKMSIGCOAVGCDROLGSKNEDKNG
VCAGDSTCLRVGQSKSHVPEKREENVIAVPLGRSVRITVKGPAHLFIESKTLQG
SKGEHFNPGVFEVNTTVEFGQSERQTFKIPGLMADFIPKTRYTAADKSVQFF
FYQISHQNRQTFDFFCTVTCGGYOLNSAECVDIRLKVVPDHYHYYPENVKPKP
LKECSMDPCSPDSGRFKEIMPDHFQPLRWEHNPWTACSVSCGGIORPSFVCEESM
HGEILQYENKMYAPKPKVMQCNLFDCPKWIAEWSCQTVTCRGLRYRVYLCINH
RGEHVGCNPQLKHIKEBCVPIPCYKPKESPVBAKLPLWAKOALQRETRITREP
TIPEPWSASTTGPQGVREYVKRVLLFTQTETELPEECEEGLTERPCLILEA
CDEPASRELDLPEDSETTWMEYAGFTPTCTATCGVGHQEAIAVCLHQTQTVND
SLCDMVRPAMSAQCTEPCPPRHWGWSGCSATCGVGIOTRVYDLPBGFTAPP
ECCRDKPHALACNDFCPGWIIHEWQCSRTCGSGTGNRRVTCROLLTQDGSFLNL
SELCQGRKASHKSCARTDCPHLAVGWSKCSVSCGVGIQRKQVQCRQARLAKGRI
PUEMCRDLPGPLVRSCQPECSKIKSEMTKIGEGQPIILSVQVRYIOTREERI
NLTIGSRVLLPNTSVIILKCPVRFKQSLIOWEKGRCLONSKRLGITKSGSLKIHGL
AAPDGVICVAGSAGDVTYVLIKIGFDNRLIARPALREPMREYPMGDHSEANLGVTW
HKRMWNKNDLIYDDHLSQNFLLRLLHGSNSAGSTNSWELKNDFEAAVKQGA
YSMDTAQDELINMSQMLTGEVSDLASOLLYQVLAELAKAQPTIMQWRGQIETP
PAAQLRGFTSVSQSHAKNSGKLTFKPKGPVLMRQSPQSPISFNKTIINSRIGNTVYI
TKRTEVINILDLITSEATYITWKDGLIOPSVKIILDTGKIQLQNPTRKQGIYE
CSVANHLGSDVSSVLYIAPVLYSVERNITKPEHNLSVWVGVEALGANVTIR
CPKVPQPNITWLRKGSISGNVSLIFNGSLILLQNVLSLENFYVCIAINALKAVA
TSVFLHLRRWPESRIVFLQGHKKYILQATNRTNSNDPTGEPPQEPPEWEPGNWHSQ
SATCGHILGARIOPCVMANGQVESEALCDHLQPLAGPEPCNIRDCPARWFTSVWSQ
CSVSCGEGVHSROVTKRTKANGTVQVSPRACAPKDRPLGRKPCFSGHPCVQVBCNR
CPGRCMGRVAVQOORTACQHSNDSNCDNRKPTLRNCTSGACDVCHWTGPKPCT
AACGRGFQSRKDCIHTRSCKPVAKRHCVOKKAPISWRHCLGPSCDRDCDITDTHYCMF
VKHLNLCSLDRYKORCCQSQEG"

source

1..7260

CDS

/organism="Homo sapiens"

/db_xref="taxon:9606"

136..5211

/note="unnamed protein product"

/codon_start=1

/protein_id="CAC69812.1"

/db_xref="GI:15593890"

/translation="MASWTSPWMLIGVMHSPLOTTAETKSPGAYFLPERALSPQG
SLEDOTIGQFLTYRYDDQTSRNRSDKDNWDAGWSDSRTGGGASYSLRRC
LGRNCEGNIRYKTKSNIDCPDAEDFRAQCSAYNDVQYQHYEYEWLPRYNDPAP
CALKHAQGNLWELAPKLVSDTRCNDSLDKMSIGCOAVGCDROLGSKNEDKNG
VCAGDSTCLRVGQSKSHVPEKREENVIAVPLGRSVRITVKGPAHLFIESKTLQG
SKGEHFNPGVFEVNTTVEFGQSERQTFKIPGLMADFIPKTRYTAADKSVQFF
FYQISHQNRQTFDFFCTVTCGGYOLNSAECVDIRLKVVPDHYHYYPENVKPKP
LKECSMDPCSPDSGRFKEIMPDHFQPLRWEHNPWTACSVSCGGIORPSFVCEESM
HGEILQYENKMYAPKPKVMQCNLFDCPKWIAEWSCQTVTCRGLRYRVYLCINH
RGEHVGCNPQLKHIKEBCVPIPCYKPKESPVBAKLPLWAKOALQRETRITREP
TIPEPWSASTTGPQGVREYVKRVLLFTQTETELPEECEEGLTERPCLILEA
CDEPASRELDLPEDSETTWMEYAGFTPTCTATCGVGHQEAIAVCLHQTQTVND
SLCDMVRPAMSAQCTEPCPPRHWGWSGCSATCGVGIOTRVYDLPBGFTAPP
ECCRDKPHALACNDFCPGWIIHEWQCSRTCGSGTGNRRVTCROLLTQDGSFLNL
SELCQGRKASHKSCARTDCPHLAVGWSKCSVSCGVGIQRKQVQCRQARLAKGRI
PUEMCRDLPGPLVRSCQPECSKIKSEMTKIGEGQPIILSVQVRYIOTREERI
NLTIGSRVLLPNTSVIILKCPVRFKQSLIOWEKGRCLONSKRLGITKSGSLKIHGL
AAPDGVICVAGSAGDVTYVLIKIGFDNRLIARPALREPMREYPMGDHSEANLGVTW
HKRMWNKNDLIYDDHLSQNFLLRLLHGSNSAGSTNSWELKNDFEAAVKQGA
YSMDTAQDELINMSQMLTGEVSDLASOLLYQVLAELAKAQPTIMQWRGQIETP
PAAQLRGFTSVSQSHAKNSGKLTFKPKGPVLMRQSPQSPISFNKTIINSRIGNTVYI
TKRTEVINILDLITSEATYITWKDGLIOPSVKIILDTGKIQLQNPTRKQGIYE
CSVANHLGSDVSSVLYIAPVLYSVERNITKPEHNLSVWVGVEALGANVTIR
CPKVPQPNITWLRKGSISGNVSLIFNGSLILLQNVLSLENFYVCIAINALKAVA
TSVFLHLRRWPESRIVFLQGHKKYILQATNRTNSNDPTGEPPQEPPEWEPGNWHSQ
SATCGHILGARIOPCVMANGQVESEALCDHLQPLAGPEPCNIRDCPARWFTSVWSQ
CSVSCGEGVHSROVTKRTKANGTVQVSPRACAPKDRPLGRKPCFSGHPCVQVBCNR
CPGRCMGRVAVQOORTACQHSNDSNCDNRKPTLRNCTSGACDVCHWTGPKPCT
AACGRGFQSRKDCIHTRSCKPVAKRHCVOKKAPISWRHCLGPSCDRDCDITDTHYCMF
VKHLNLCSLDRYKORCCQSQEG"

BASE COUNT 2022 a 1593 c 1811 g 1834 t

ORIGIN

Query Match 19.3%; Score 1023.4; DB 6; Length 7260;
Best Local Similarity 60.9%; Pred. No. 2.3e-173;
Matches 1744; Conservative 0; Mismatches 1076; Indels 45; Gaps 3;
QY 60 cctgagttccagaccacacgtccgagagagcggagcgccctatggaatgctggg 119
DB 321 CCAGACTCAAGAAACACTCGTTCCAGATGAAGACAAAGATGGCAACTGGGATGCTTGGGG 380
QY 120 cccatgagtgaaatgctcacgcacctcgqgggtggggccctcctactctctgagggcgctg 179
DB 381 CGACTGGAGTGACTGCTCCCGGACCTGTGGGGGAGGAGCATCATATCTCTCGGGAGATG 440
QY 180 cctgagcagcagagctgtgaaggaagaataatccgatatagacaacatgcagtaagtgtgga 239
DB 441 TTTGACTGGAGGAATTTGTAANGGCGACACATTCGGGTACAAGACATCGACCAATCATGA 500
QY 240 ctgcccaccagaacagtgtaattccagctcagcaatgtcagctcataatgatgtcaa 299
DB 501 CTGCCCCCCAGATGAAGAGATTTTCAGAGCCCGCAGCGTGTCTCAGCCTACAAATGATGCCA 560

QY	1380	ggctctctgcatcgaccatcgaggaaatgcacacagagaggtgtgtgccccaaaaaagaagcc	1433
Db	1641	TGTTCTGTGTATTAAACCCACGGGAGAGCATGTTGGGGCTGCAATCCACAACCTGAAGTT	1700
QY	1440	ccacataaagaggaatgcatacctaccactccctgctataaaccccaagagaaacttcc	1499
Db	1701	ACATCTCAAGAAGAATGTGTCATTCCTCCATCCCGTGTATTAAACCAAAAGAAAAGTCC	1760
QY	1500	agtcgagggccaagtgtccattggtttcaaacaaactcaagactgaagaaggagtgctgt	1559
Db	1761	AGTGAAGCAAAATTCCTTGCTGNAACAGCACAGAACTAGAGAGACCAGATAGC	1820
QY	1560	gtcagagaggccctcgtttcatccacagagcctggtgcgcctgcacagtccactgtgtgt	1619
Db	1821	AACAGAGAACC AACCTTCATTCCAGAACCCCTGGTCAGCCTGCAGTACCACGTGTGGGCC	1880
QY	1620	ggggaccagggtcgaataagtccagtgccagggtgctcctgtttctcagtcctgtgc	1679
Db	1881	GGGTGTGCAGGTCCGTGAGGTGAAGTGGCGGTGTGCTCTCAATTCACGCAGACTGAGAC	1940
QY	1680	tgacctgcctattgacagagtgtgaaggggcccaagccagcatccccagcgtgccgtgtatgc	1739
Db	1941	TGAGCTGCCCGAGGAAGAGTGTGAAGGCCCAACAGTGCACCAC-----GA	1985
QY	1740	agggccatgcagcggggaaattcctgagtccaacccagacagacagatgggtctttgg	1799
Db	1986	ACGGCCCTGCCTCTCTGGAAGCATGTGATGAGAGCGCGCCCTCCCGAGAGCTAGACATCCC	2045
QY	1800	tggcctgcaggatttcagcagcgtatgactggaggtatgaggggttaccaaagtgcctc	1859
Db	2046	TCYCCCTGAGGACAGTGAACGACATTCAGCTGGGAGTACGCTGGGTTCACCCCTTGCAC	2105
QY	1860	c9agtcctgtggaggaggtgtccagagagcgtgtgtgagctgcttgacaaacagactcg	1919
Db	2106	AGCAACATGCTGGGAGGCCATCAGAGGCCATAGCAGTGTGCTTACATATCAGACCCCA	2165
QY	1920	ggagcctgtgtaggagaacctgtgctgaccagccgcgcgcgcccaagctcctgaagtc	1979
Db	2166	GCAGACAGTCAATGACAGCTTGTGTGATATGGTCCACCGTCTCCAGCCATGAGCCAGGC	2225
QY	1980	ctcaatttggatccctgcceagaaggtggaaattggcaagtggagttcagtgcattgtagtct	2039
Db	2226	CTGTACACAGAGCCCTGTCCCCCAAGTGGCATGTGGGCTCTTGGGGGCCCTGCTCAGC	2285
QY	2040	cacatgtggggtcggctacagaccagacagcgtctcttgacgcacactgttctccagaga	2099
Db	2286	TACCTGTGGAGTTGGAAATTCAGACCCGAGATGTGTACTGCTGCACCC-----2332	2332
QY	2100	gatgaatgaacagtcactcgtggtgtgagtgtgtccgcagcccaagcccagcacggt	2159
Db	2333	-----CAGGGGAGAGCCCTCTCTCTCTGAGAGTGCAGATGAAAAGGCCCATGCTTT	2387
QY	2160	gcaagctgtaccgcttattgccccccagcctggtaccctgcacagtggcagccgtg	2219
Db	2388	ACAAGCATGCAATACAGTTTACTGCCCCCTCTGGCTGGCAATTGAAGAAATGGCAGCAGTG	2447
QY	2220	ttccagaacggtgcccgggggtgttcagaacacgtgaggttctttgcaagcagcgcatgpc	2279
Db	2448	TTCCAGGACTTGTGCGGGGAACTCAGAACAGAAAGTAGTCACTGTGCGGACGTGCTAAC	2507
QY	2280	tgatggcagcttctcgtggagcttcctgngacctctgttcaagttccaacacctgcctgcca	2339
Db	2508	GGATGCAAGCTTTTGAATCTCTCAGATGAATTTGTGCCAAGGACCCCAAGCATCGTCTCA	2567
QY	2340	gcaagcatcgaagaaagatgactgtccacgcagtggtctctctcagactggacagagtgc	2399
Db	2568	CAAGTCTGTGCCAGACAGACTGTCTCTCCACATTTAGCTGTGGGAGACTGTGTCGAAGTG	2627
QY	2400	ttccacaagctgcgggggaagccaccagacttcgaagccatttgcgcgaagaagtgcagaa	2459
Db	2628	TTCTGTCAAGTGTGGGTGGTAATCCAGAGAAGAAACAGAGTGTGTCAAGGCTGCACGC	2687
QY	2460	aaccggcctctcaacggtgtgcaattccaacctgtgcccgcacctgcttcttctctcctc	2519

RESULTS

```

AX224809.1
LOCUS       AX224809                5076 bp    linear    PAT 10-SEP-2001
DEFINITION  Sequence 1 from Patent WO0161011.
ACCESSION   AX224809
VERSION     AX224809.1  GI:15554909
KEYWORDS    .
SOURCE      human.
ORGANISM    Homo sapiens
REFERENCE   1 (bases 1 to 5076)
AUTHORS     Donoho,G., Scoville,J., Turner,C.A., Friedrich,G., Zambrowicz,B.,
            and Sands,A.T.
TITLE       Novel human thrombospondin repeat proteins and polynucleotides
            encoding the same
JOURNAL     Patent: WO 0161011-A 1 23-AUG-2001;
            Lexicon Genetics Incorporated (US)
FEATURES    Location/Qualifiers
             1..5076
             /organism="Homo sapiens"
             /db_xref="taxon:9606"
BASE COUNT  1384 a 1227 c 1338 g 1127 t
ORIGIN

```

	Query Match	19.3%;	Score 1020.2;	DB 6;	Length 5076;
	Best Local Similarity	60.8%;	Pred. No. 8.9e-173;		
	Matches 1742;	Conservative	0;	Mismatches 1078;	Indels 45; Gaps 3;
QY	60	cctgaggtccaggaacgcgcacgcctccgaggaagaccgggaacggcctatggagatgcctgagg	119		
Db	186	CCAGACCTCAAGAAACACTCGTTCAGATGAAGACAAAGATGGCAACTGGGATCGTTGGGG	245		
QY	120	cccatggagtgaatgctccacgcacactcgcgggggtggggcctctactctctgagggcgtg	179		
Db	246	CGACTGGAGTGACTGCTCCCGACCCTGTGGGGGNGGACATCATATCTCTCGCGAGATG	305		
QY	180	cctgagcagcaagactgtgtaaggaagaaaaataccgatacacagaacalcgagtaatgga	239		
Db	306	TTTGACTGGAAGGAANTGTGAAGGGCAGAACATTCGGTACACACATGACGAATCATGA	365		

Qy	240	ctgcccaccgaagcagggtgatttccgagctcagcaatgctcagctcataaigtatgtaaa	299
Db	366	CTGCCTCCAGATGACAGAAGATTTCAGAGCCAGCAGTGCTCAGCCTACAATGATGTCCA	425
Qy	300	gcacatggccagtttataagatgcttctgtgctaaagaccctgacaaacccatgttc	359
Db	426	GTATCAGGGGATTAATAATGAAGCTTCCACGATATAATGATCTGTGTCGCCGTGTC	485
Qy	360	actcaagtgcgaagcgaaggacaacccctggtgttgaaactagcacctaaagtctctaga	419
Db	486	ACTCAAGTGTATGCACAHGGACAAAACCTGGTGTGGNGCTGGCACCTAAGGTACTGA	545
Qy	420	tggtagcgttgctatcacagaactcttggatatgtgcatacagtggtttatgccaattgt	479
Db	546	TGGAACCTGTTGCAACACGGACTCTTGACATGTGTATCAGTGCACTGTCTCAGCAGT	605
Qy	480	tggctgcgataccacagctgggaagcacgctcaagaaataactatgggtgctgcacag	539
Db	606	GGGCTGCCATCGGCACCTCGGNAAGCANTGCCAHGAGGACAACTGTGGAGTCTGTGCCGG	665
Qy	540	agatgggtccacctcgcgctgggtccgagggcagttataaatcccagctctccgcaccaa	599
Db	666	CGATGGCTCCACCTCAGGCTGTACGGGGACAATCAAAAGTCACACGTTTCTCTGAAAA	725
Qy	600	atcggatgatactgtggttgcaattccctatgcaatgaaatagacatactgccttgtctaaa	659
Db	726	AAGAAGAAATGTAATGTCTGTTCCTTTGGGAAGTCGAAAGTGCAGAAATTCACGTGA	785
Qy	660	aggtcctgatcacttatacttgaaaccaaaccctccaggggactaaaggtgaaacacag	719
Db	786	AGGACCTGCCACCCTCTTATTGAATCAAAAACACTTCAAGGAAGCAAGGAGACACAG	845
Qy	720	tctcagctccacaggaaatttctctgtggacaattctagtgggacttccagaaatttcc	779
Db	846	CTTTAACAGCCCCGCGCTTTTGTGCTAGAAAACACACAGGTGAATTTTCAGAGGGGCTC	905
Qy	780	agacaaagagatactgagaatgctgtgaccactccacagcagatttcattgtcaagattcg	839
Db	906	CGAGAGGCCAAACTTTTACATTTCCAGACCTCTGTATGGGTGATTTCTATCTTCAAGACCAG	965
Qy	840	taactcgggtctccgtgcaggtacagttccagttccattcttctatcaacccatcatccacg	899
Db	966	GTACACTGCAGCCAAAGACAGCGTGGTTCAGTTCTCTTTTACCAGCCCATCAGTCATCA	1025
Qy	900	atggaaggagacggaatttcttctgtctcagcaacctgtgagagaggttaacagctgac	959
Db	1026	TTGGAGACAAACTGACTCTTTTCCCTGCAGCTGTGACGTGTGAGGAGGTTTATCAGCTCAA	1085
Qy	960	atcggctgagtgctacgatactgaggacaaccgtgtgggtgctgaccaatactgtcacta	1019
Db	1086	TTCTCTGAAATGTGAGATATCCGCTTGAAGAGGTAGTTCCTGACCANTATTGTGCATA	1145
Qy	1020	ttaccagagaaacatacaacccaaacccaaagcttcaggagtgcaacttggactcttctgtcc	1079
Db	1146	CTACCTGAAATGTAAAACCAAAACCAAACTGAAGGAATGCAGCATGGATCCCTGCC	1205
Qy	1080	agcgaagtgcgatatcaagcagatcatgcttatgacctctaccatcccccttctcgggtg	1139
Db	1206	ATCAAGTGAATGATTAAAGAGATAATGCCCTATGACCACCTTCCAACTCTTCTCGCTG	1265
Qy	1140	ggagggccacccatggaccgctgtctcctcctcgtgtgggggggcataccagagccgggc	1199
Db	1266	GAAACATACTCTGGACTGTCATGTTCCTGTCTGTGGAGGAGGATTCAGACACGGAG	1325
Qy	1200	agttctcgtgtggaggaggaatccacaggggcagtgctcacttcagtggaagagtggaaatg	1259
Db	1326	CTTTGTGTGTAGAGGAATCCATGCATGGAGAGATATTGCAGGTGGAAGAATGGAAGTG	1385
Qy	1260	catgtacaccccttaagatgcccatcgcagccctgcacaacttttgcctgcctaaatg	1319
Db	1386	CATGTACGCCCAACCAACCAAGGTATGCAAACTTCTAATCTGTTGATGTCGCCCAAGTG	1445

Qy	1320	gctggccaggagtggtctcctgtgcacagtcagatgtgtgcccaggcgctccagataccctgtt	1379
Db	1446	GATTGCCATGGAGTGGTGCTCAGTCGACACTGACTGTCTTGCCCGCAGGGTTTACGGTTACC	1505
Qy	1380	ggctcctgatcagcaccatcgaggaatgcacacagagagctgtagccccaaaaacaagaacc	1439
Db	1506	TGTTCTGTGTAATTAAACCACCGGAGAGCATGTTGGGGGCTGCATACTCACAACTGAAGTT	1565
Qy	1440	ccacataaaagagaattgatgtacccactccctgtcataaacccaaaagagaaccttcc	1499
Db	1566	ACACATCAAGAGAATGTGTCAITCCCATCCGTGGTGTATTAACCAAAAAGAAAAAGTCC	1625
Qy	1500	agtcaagccaaattgccattggtttcaaacaaagctcaaagactagaagagagctgtgt	1559
Db	1626	AGTGAAGCAAATTTGCTTGGCTGAACAACAGCAAGAACCTAGAAGAGACAGCAATAGC	1685
Qy	1560	gtcaagagagcccttgttcaatccagagagcctggctggcctgcacagtcacactgtgtgt	1619
Db	1686	AACAGAAGAACAACGTTCAITCCAGAACCCTGGTCAGCTCAGTACACGTTGGGGCC	1745
Qy	1620	ggggaccacaggtgcgaatagtaggtgcaggtgtcctgtctcttctctaatgcctgtgc	1679
Db	1746	AGGTTGTCAGSTCCGGAGGTGAAGTGCCTGTGCTCTCATATTCACGCACATGAGAC	1805
Qy	1680	tgaacctgctattgaacagtgtaagggcccaagccagcatcccaagctgctgttatgc	1739
Db	1806	TGAGCTCCCGCAGAGAGAGTGTGAAGCGCCCCAAGCTGCCACG-----GA	1850
Qy	1740	aggcccatcagcggggaaattcctgagttcaaccacagacagacagatgggctctttgg	1799
Db	1851	ACGGCTTGCTCTGTGNAACATGTGATGAGAGCCCGGCTTCCGAGAGCTTAGACATGCC	1910
Qy	1800	tggcctcaggatttcagcagcgtgtatgactggagtagtaggggtttccaacagtgctc	1859
Db	1911	TCTCCCTCAGSACAGTCAGACGACTTACGACTGGGAGTACGCTGGGTTCACCCCTTGCA	1970
Qy	1860	cgaatcctgtggaggaagtgltccagagagcgtgtgtgtagctgtcttgaacaaacagactcg	1919
Db	1971	ASCAACATGCTTGGGAGGCCATCAAGAAGCCCATAGCAGTGTCTTACATATCCAGACCCA	2030
Qy	1920	ggagcctgctgaggaagaacctgtgcctgaccagcgcgcgcgcccaagctcctgaagtc	1979
Db	2031	GCACACAGTCAATGACAGCTTGTTGATATGTFPCACCGCTCTCCAGCCATGAGCCAGGC	2090
Qy	1980	ctgcaatttgatccctgcccagcaaggtgggaaattggcaagtgtagtccatgtagctct	2039
Db	2091	CTGTAAACACAGAGCCCTGTCCCCCGAGGTGGCATGTGGGCTCTTGGGGCCCTGCTCAGC	2150
Qy	2040	cacatgtggggtgcgcclacagaccagagacgtctctctgcagccacctgcttccagaga	2099
Db	2151	TACCTGTGGAGTTTGGAAATTCAGACCGGAGATGTGTACTGCCPWCACC-----	2197
Qy	2100	gatgaatgaacagtcatectgctgtatagctgtgtgcagccagccagccagcacggt	2159
Db	2198	-----CAGGGGAGNCCCTGCCCTCTGAGGAGTGGCAGATGAANAAGCCCATGCTTT	2252
Qy	2160	gcaagctgttaacgcgtttaattgccccccagcctggttacctctgcacagtgccagccgtg	2219
Db	2253	ACAAAGCATGCAATCAGTTTTCAGTGCCTCTCTGGCTGGCACATTTGAAGAATGGCAGCATG	2312
Qy	2220	ttccagaacgtgtgcccgggggtgttcagaaaacgtgaggttctttgcagacgacgcalggc	2279
Db	2313	TTCCAGGAGCTTGTGGCGGGGAACTCAGAACAGAAAGAGTCACTGTGCGGACGTGCTAAC	2372
Qy	2280	tgatgcagcttctgagatttctgagaccttctgttcagcttcaaacctgcctgcaca	2339
Db	2373	GGATGGCAGCTTTTTTGAATCTCTCATGAATTTGTGCCAAGGACCCCAAGGCATCGTCTCA	2432
Qy	2340	gcaagcatgcagaagaatdacigtctccagcagtggtcttctcagactggacagatgt	2399
Db	2433	CNAGTCTGTGCCAGGACAGCTGCTCTCCACATTTAGCTGTGGGAGACTGGTGTGAAGTG	2492
Qy	2400	ttccaacagctgcggggaaagccacagactcgaagcgcatttgtccaaaagatatcttaa	2459


```
QY 1388 gcatcgaccatcagagggaatgcacacagagagctgtagcccaaaaaaagcccccacataa 1447
Db 1292 GTATTAAACCAACCGGAGAGATGTTGGGGCTGCAATCCACAACCTGAAGTTACACATCA 1351
QY 1448 aagagaatgcacgtaccacactccctgtctataaacccaaagagaaaacttccagtgagg 1507
Db 1352 AAGANAATGTGCTATCCCATCCCGTGTATTAAACCAAAAGAAAAGTCCAGTGGAAAG 1411
QY 1508 ccaagttggcatgtttcaaaagaactcaagagctagagaagagagctgtgtgtcagagg 1567
Db 1412 CAAAAATGGCTTGGCTGAACAGCACAAGAAGCTAGAGAGACCCAGAAATGACACACAAG 1471
QY 1568 agccctgttcatcccaagagccctgtgtccctgcacagtcacactgtgtgtgtggggaccc 1627
Db 1472 AACCAAGCTTCATTCACAGAACCCTGTGTGAGCTTACCAACGCTGTGGCCAGGTGTGC 1531
QY 1628 agtgcgaatagtcaggtgcaggtgcctgtctctctctctctctctctctctctctctct 1687
Db 1532 AGGTCCGAGGTGAAGTGCCTGTCTCTCATTTACAGCAGACTGAGACTGAGCTGC 1591
QY 1688 ctattgacgagtgtagaggggcccaagccagcgcacatcccgagctgtgtgttatgcaggccat 1747
Db 1592 CCGAGGAAGAGTGTGAAGGCCCAAGCTGCCACC-----GAACGGGCCCT 1636
QY 1748 gcagcggggaaattccttgagttcaaacccagacagagacagatgggctcttgtgtgcctgc 1807
Db 1637 GCCTCTCGAAGCATGTGTATGATGAGAGCCCGCCCTCCCGAGAGCTAGACATCCCTCTCCCTG 1696
QY 1808 agtatctcgacgagctgtatgactggaggtatgaggtttcccaagtgtctcgagttcct 1867
Db 1697 AGNACGTGAGACACTTACGCTGGGAGTACGCTGGGTTCACCCCTTGCACAGCAACAT 1756
QY 1868 gtgagagagtggtccagagagctgtgtgtgagctgtgtgaacaaacagactcgggagcctg 1927
Db 1757 GCTTGGGAGGCCATCAAGAAGCCATACAGTGTGCTTATACATATCCACACCAGCAGACAG 1816
QY 1928 ctgaggagaacctgtgcgtgaccagccgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 1987
Db 1817 TCAATGACAGACTTGTGTGTATGTGTCCACCGCTCTCCAGGCATGAGCCAGGCCGTGAACA 1876
QY 1988 tggatccctccagacaagtggaattggcaagtggatgcaatgagttcactatgctcactgtg 2047
Db 1877 CAGAGCCCTGTCCCCAGGTGGCATGTGGGCTTGTGGGGCCCTGTCTACGTAACCTGTG 1936
QY 2048 ggtcgtgcctcacagacagagactcttctgcagccacactgtcttccagagagatgaatg 2107
Db 1937 GAGTTGGAAATCAGACCCGAGAGTGTGTACTGCCTGCACCCAGGGGAGAC----- 1985
QY 2108 aaacagtcatccttgctgtgagctgtgtgcgcagcccaagcccaagcagcaggtgcaagctt 2167
Db 1986 -----CCCTGCCCTTCCCTGAGGAGTGCAGATGAAGAAAAGCCCATGCTTTACAAGCAT 2038
QY 2168 gtaacgctttaattgccccccagcctgttaccctgcacagtggcagcctgtttccagaa 2227
Db 2039 GCATCAGTTTGTACTGCCCTCTCTGGCTGGCCACATTAAGAAATGGCAGCTGTTCAGGA 2098
QY 2228 cgtgtggcgggggtgttcagaanaactgtaggttcttcttgcagagcgcagtcgctgtagtgc 2287
Db 2099 CTGTGTGGGGGGAACCTCAGACAGAAAGATGCACCTGTCTGCGAGCTGCTAACGGATGCGCA 2158
QY 2288 gcttctggagcttcttgaaactctgttcagcttcaaacctgcctgcagcaagcat 2347
Db 2159 GCTTTTGAATCTCTCAGATGAATGTGTGCCAAGCACCACCAAGGCATGCTCTCACAAAGTCT 2218
QY 2348 gcaagaaagatgactgtcccgagagtggtcttctctcagactgagacagagtgtttccaaa 2407
Db 2219 GTGCCAGGACAGACTGTCTCCACATTTAGCTGTGGGAGACTGTGTGAAGTGTCTGTCA 2278
QY 2408 gctgcgggggaagccacacagactcgaagcgcattttgccgaagatgctgaaacccgccc 2467
Db 2279 GTTGTGTGTGTGAATCCACAGAAAGCAGGCTGTCTCAAGGCTGGCAGCCCAAGGTC 2338
QY 2468 tctcaacggtgttcaattccacctgtgcgcgcgcctgtcttctctctctctctctctctc 2527
```

```
Db 2339 GGCGCATCCCTCAGTGCAGATGATGTGCAGGATCTACCAGGTTCCCTCTCTTGAAGAT 2398
QY 2528 cctgtatgtctg-----aacctgtgaagggcccgggcgccatccacgaagc 2575
Db 2399 CTTGCCAGATCCCTGAGTGCAGTAAATCAATCAGAGATGAAGACAAACTTGGTGAGC 2458
QY 2576 acagccgcacatcgcgccgcccaggaaggtctacatccacactcgagcgaggaagc 2635
Db 2459 AGGTCCGCAGATCTCAGTGTCCAGAGAGTCTACATTCAGACAAGGAAGAGCGTA 2518
QY 2636 tgcacttctgtgtggggggtctcgcctacactgtctcccaagacggcggtgtgtgcgct 2695
Db 2519 TTAACCTGACCATGTTAGCAGAGAGCCTATTTGCTGCCCAACACATCCCTGATTATTAAGT 2578
QY 2696 gccggcgcgagaggtccgcaagccctcatcactcgtggaggaagcgccagcacctca 2755
Db 2579 GCCCGCTGCAGCATTCAGAAATCTCTGATCCAGTGTGCACAGGAAAGATGCCGTTGCCTGC 2638
QY 2756 tcagctgcagcgacgtcacggtggcccttgcgtctatctcaagatccacgcctcaagc 2815
Db 2639 AGAATCCAAACGCTTGGCATCACAAGTCAGGCTCACTAAAAATCCACCGTCTTGGCTG 2698
QY 2816 cctcgagatgcaggggtctacacctgtcagcgggcccgggcgagcactttgtgatta 2875
Db 2699 CCCCAGCATCGCGTGTACCGGTGCATTGCAGCTCTGCACAGGAAACAGTTGTGCTCA 2758
QY 2876 agtcatcgtgaggaacacgcaagctcgtggcccgcc 2912
Db 2759 AGCTATTGTTACTGACAAACCGGCTCATCGCACGCC 2795

RESULT 8
AX224819 LOCUS AX224819 2538 bp DNA linear PAT 10-SEP-2001
DEFINITION Sequence 11 from Patent WO0161011.
ACCESSION AX224819
VERSION AX224819.1 GI:15554914
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 2538)
AUTHORS Donoho, G., Scoville, J., Turner, C.A., Friedrich, G., Zambrowicz, B.
and Sands, A.F.
TITLE Novel human thrombospondin repeat proteins and polynucleotides
encoding the same
JOURNAL Patent: WO 0161011-A 11 23-AUG-2001;
Lexicon Genetics Incorporated (US)
FEATURES
source 1..2538
/organism="Homo sapiens"
/db_xref="taxon:9606"
BASE COUNT 670 a 627 c 678 g 563 t
ORIGIN

Query Match 17.6%; Score 930.4; DB 6; Length 2538;
Best Local Similarity 63.3%; Pred. No. 1.1e-156;
Matches 1479; Conservative 0; Mismatches 826; Indels 33; Gaps 2;

QY 60 cctgagttccagaccgcagctccgagagagcagcgagcgcctatgggatcctgggg 119
Db 186 CCAGACCTCAAGAAACACTCGTTCAGATGAAGACAAGATGGCAACTGGATGCTGGGG 245
QY 120 cccatgagtgaaatgctcacgcacctgtcgggggtggggcctcctactctctgagcgctg 179
Db 246 CGACTGGAGTGAAGTCTCCGACCTGTGGGGGAGGAGCAICATATTTCTTCGCGGATG 305
QY 180 cctgagcagcaagagctgtgaaggaagaataatccgatacagaacatgcaatgtgga 239
Db 306 TTTGACTGGAAGGAATTTGTGAAGGGCAGACATTCGGCTACAGACATGCAGCAATCATGA 365
```


RESULT

AF237652	2523 bp	linear	PRI 02-MAR-2001
LOCUS	Homo sapiens a disintegrin-like and metalloprotease domain with thrombospondin type I motifs-like 3 (ADAMTSL3) mRNA, partial cds.		
DEFINITION	AF237652		
ACCESSION	AF237652.1		
VERSION	GI:13183077		
KEYWORDS	human.		
SOURCE	Homo sapiens		
ORGANISM	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.		
REFERENCE	1 (bases 1 to 2523)		
AUTHORS	Hirohata, S., Anand-Apte, B., Seldin, M. and Apte, S.		
TITLE	Punctin, a member of a new family with similarities to ADAM-TS proteases, is a component of extracellular matrix of skeletal muscle		
JOURNAL	Unpublished		
REFERENCE	2 (bases 1 to 2523)		
AUTHORS	Anand-Apte, B. and Apte, S.		
TITLE	Direct Submission		
JOURNAL	Submitted (21-FEB-2000) Biomedical Engineering, Cleveland Clinic Foundation, 9500 Euclid, Cleveland, OH 44195, USA		
FEATURES	Location/Qualifiers		
	1..2523		
	source		

```

/organism="Homo sapiens"
/db_xref="taxon:9606"
1. .>2523
/gene="ADAMTSL3"
225. .>2523
/gene="ADAMTSL3"
/note="ADAM-TSL3; similar to GenBank Accession Number
AB033059; extracellular matrix protein"
/codon_start=1
/product="a disintegrin-like and metalloprotease domain
with thrombospondin type I motifs-like 3"
/protein_id="AAK15041.1"
/db_xref="GI:13183078"
/translation="MASWTSPPWWLVIGMFVMSPIQTPTAERKSPGAVLPEFALSPQG
SFLEDTQEQFLTYYDDTSTRTSDEKDGNDWAGDWSDCSTCGGASYSLRRC
LTGRNCEGNIYKTCNSHDCPPDAEDFRAQCQSAYNVQYGHYEWLPRTNDPAAP
CALKHAQGNOLYVELAPKYLVDGTRCNDLSLDCISGICQACDQRQGLSNKEDNCG
VCAGDSGICRLVRGQSGSHVPEKREENVIAVLGSRSVRIYTKGVPHLFIESKTIQK
SKGEHSFVGVFVNTTVEFGSRSEKTEQIFKPLMADFIKTRYTYAAKSDVVOQF
FYQTSLSHQWRQDTFFPCTYCGGGYQLNASECVDRLKRVDPHDYCHYYFVNVKPKPK
LKECSDMSPGSDGKELMPYDHPFLPRWEHNPTACSVSCGGIORESFYCVVEESM
HGEILQVEWKMVAPKPKVMOTCNLFDCPKWIMWASQCTYTCRGLRKYRVVLCLNI
RGEHVGCGNPOLKHIKECVYIPICYKPKESPVBAKLPLWIKQAQELSEATIEEP
TFIPEPWSACSTTCGPGVQVREVRKCVLLFTQTETLGHQEECPGLPTERPCLEEA
CDSPASRELDLIPDESEDTYDWEYATFPCTATCLGHOEAIACLVIHQQTQVND
SLCDMVKHPPPAMSAQCNETPCCPPRHWVHSGWSGPCSATCGVGIOTRDVYCLHPGETPAP
EECDREKPHALQACNQFCPCPGWHIE"
635 a 652 c 692 g 544 t

```

Query Match 16.6%; Score 878.2; DB 9; Length 2523;
Best Local Similarity 65.1%; Pred. No. 2.5e-147;
Matches 1319; Conservative 0; Mismatches 693; Indels 15; Gaps 1;

[illegible]

D	b		590	CTGCCCTCCAGATGCGAAGAATTTCAGAGCCCACGAGTGCCTCAGCCCTACAATGATGTCCA	649
Q	y		300	gcaccatggccagcttttatgaattggttctcctgttgtctaataagacctcgacaaccccattgttc	359
D	b		650	GTAFCAGGGCAATTACTATGAATGGGTTCACCAGATAAATGATCCCTGTGCCCCGTGTGC	709
Q	y		360	actcaagtgcacaaggcaaaacccctgggttgtaactagacataaagttcttaga	419
D	b		710	ACTCAAGTGTCATGCACAGGACAAACCTTGGTGGTGAGCTGGCACCTTAAGGTACTGGA	769
Q	y		420	tggtaoagctgtgatatacacagaactcttggatatgtgcatacagtggttttatgccaaattgt	479
D	b		770	TGGAACCTCGTTGCAACACGACACTCTTGACAATGTGTATCAGTGGCATCTGTCAGGCAGT	829
Q	y		480	tggctcgatcaccaactgggaagacccgtcaaggaagataaacttgagggtctgcaacgg	539
D	b		830	GGGCTGGGATTCGGCACTGGGAAGCAATGCCAAGGAGGACAACTGTGGAGTCTGTGCCGG	889
Q	y		540	agatgggttcacactgcgcggctggtccgagggcagtataaaaccacgtctccgcgaacca	599
D	b		890	CGATGGCTCCACCTGCAGGCTTGTCGGGGACAATCAAGTCACACGTTCTCTCTGAAAA	949
Q	y		600	atcggtatgatactgtggttgcaattccctatgtggaagtagacatatcgccttgtcttaaa	659
D	b		950	AAGAGAGAAAATGTAAATTTGCTTTGCTTTGGGAAGTCGAAGTGTGAGAAATACAGTGAA	1009
Q	y		660	agtcctgtacacttatatctggaaccaaaccctccagggagctaaagttgaaacag	719
D	b		1010	AGCACCTGTCCACCTCTTATTGAAATCAAAAACACTTCAAGGAACCAAGAGGAGAACACAG	1069
Q	y		720	tctcagctccacaggaacttcttggtagacaattctagttggacttccagaaattccc	779
D	b		1070	CTTTAACAGCCCCGGCGCTTTGTGCTGTA AAAACACAACAGTGGAAATTCAGAGGGGCTC	1129
Q	y		780	agacaagaatactagagaatggctggaccactcacagcagatctcatctgtctcagattcg	839
D	b		1130	CGAGAGCAAACTTTTAAGATTCCAGGACCTCTGATGGCTGATTTTCATCTTCAACACACAG	1189
Q	y		840	taactcgggtccgctgacagtlacagtcaggttcatcttctataccaacccatccaccgc	899
D	b		1190	GTACACTGCGCAAAAGACACAGCGTGGTTTCAGTTCCTTTTACCAGGCCATCAGTCATCA	1249
Q	y		900	atgagaggagacgatttcttctcctgtctcagcaacctgtggaggaggttatcagctgac	959
D	b		1250	GTGGAGACAAATGACTCTTTCCCTGTCACTGTGACGTGTGGAGGAGGTATTCAGCTCAA	1309
Q	y		960	atcggctgagtgctacgactctgaggagacaaccggtgtgtgtcagccaataactctcata	1019
D	b		1310	TTCTGCTGAATGTGTGGATATCCGCTTGAGAGGGTAGTTCCTGACCATATTATGTGCATTA	1369
Q	y		1020	ttaccagagacatcaaacccaacccaagcttcaggagtgcaacttgatcccttgttccc	1079
D	b		1370	CTACCCCTGAAATGTAA AACCAAAAACCAAACTGAAGAAATGCAGCATGGATCCCTGCC	1429
Q	y		1080	agccagtgcaggtatacaagcagatcatccttatgaaccttaccatcccccttctcgggt	1139
D	b		1430	ATCAAGTGATGATTTAAGAGATAATGCCCTTATGCCACCTTCACAACCTCTCTCTCGCTG	1489
Q	y		1140	ggaggccaccccatggaccgctgctcctcctcgtgtggggggggccatccagagccgggc	1199
D	b		1490	GGAACATAATCTTGGACTGTCATGTTCCGTGTCTGTGSGAGGGGATTCAGACCGGAG	1549
Q	y		1200	aqfittcctgtgtggaggaggacatccaggggcatalcacttcadlggaagagtggaaatg	1259
D	b		1550	CTTTGTGTGTGTAGAGGAATCCATGATGTGAGAGATATTGTCAGGTGGAAGAAATGGAAGTG	1609
Q	y		1260	catgtacaccccetaaatgccccatcgccgcagccctgcaacatttttgactgcccataatg	1319
D	b		1610	CATGTACGACCAACCAAGGTATGCAAACTTGTANTCTGTTTGTATGGCCCCCAAGTG	1669
Q	y		1320	gctggcacagdgatggtctccgtgcaacgtgacatgtgacatgtggccagggcctcagatccqgt	1379

Db 1052 ATCTTGACATGCAATGTCCTGTCGAGGAGGATTCAGAGACGGAGCTTTGTGT 1111
Qy 1208 gtgtggaggagagacatccagggcagtgctcaacttcagtggaagagtgaaatgcatgtaca 1267
Db 1112 GTGTAGAGGANTCCATGCATGGAGNATATTCAGAGTGGAGTGAAGTGCATGTACG 1171
Qy 1268 cccctaagatgcccatacgccagccctgcaacatttttgactgcccataatggctggcac 1327
Db 1172 CACCCAAACCAAGTTATGCAAACTTCTAATCTGTTGATTCGCCCAAGTGGATTGCCA 1231
Qy 1328 aggaatggtccgtgcaatgacatgacatgagccagggccctcagatataccgtgtgctcctet 1387
Db 1232 TGGAGTGTGCTCAGTGCACAGTACTTGTGGCCGAGGGTTACGGTACCGGGTTGTCTGT 1291
Qy 1388 gcatcgacatcgagggaatgcacagagagctgtgtagcccaaaaaaacacacataa 1447
Db 1292 GTATTAAACCCCGGAGAGCATGTGGGGCTGCATCCACAACTGAAGTTACATCA 1351
Qy 1448 aagaggaatgcataccactccctgctataaacccaaagagaacactccagtcgagg 1507
Db 1352 AAGAAGAAATGTCTATCCCATCCCGTGTATTAACCAAAAGAAAAAGTCCAGTGGAG 1411
Qy 1508 ccaagttgcatgttcaacaagctcaagagctagaagaagagagctgtgtcagagg 1567
Db 1412 CAAATTCCTTGCTGAAACAGCAAGAACTAGAGAGACCCAGATAGCAACAGAA 1471
Qy 1568 agccctgttccatccagagagctggtcgctgcacagtcacactggtgtgggaccc 1627
Db 1472 AACCACAGTTTCATCCAGAACCTGGTCNGCTGCAGTACCACGCTGGGCCAGGTGTC 1531
Qy 1628 aggtgcgaatagtcaggtgcccagtgctctctgttcttctcagtcogtgcactgc 1687
Db 1532 AGTCCCGAGGTGAAGTGGCGTGTCTCTCAATTCACGACACTGAGACTGAGCTGC 1591
Qy 1688 ctattgacagagtggaagggcccaagcagcatcccaagcgtgctgttatgcagcccat 1747
Db 1592 CCGAGGAGAGTGTGAAGGCCCAAGCTGCCACC-----GAACGGCCCT 1636
Qy 1748 gcagcggggaaattcctgagttcaaccagacagagacagatgggctcttgggtggcctgc 1807
Db 1637 GCCTCTCGAAGCATGTGATGAGAGCCCGCCCTCCCGAGAGCTAGACATCCCTCTCCCTG 1696
Qy 1808 aggaattcgacagactgtatgactggagtgatgaggggttcaacaagtctccagtcct 1867
Db 1697 AGCAGCTGACACGACTTAGCAGTGGGAGTACGCTGGGTTTCAACCCCTTGACACCAAT 1756
Qy 1868 gtgagagagtgctcagagagctgtgtgagctgttgaacaaacagactcggagcctg 1927
Db 1757 GCTTGGAGGCCATCAGAGAGCCATACAGTGTGCTTACATATCCAGACCCAGCAGACAG 1816
Qy 1928 ctgagagaaacctgtgctgacagcgcgcggcccccacagctcctgaagtccctgcaatt 1987
Db 1817 TCAATGACAGCTGTGTGATATGTTCCACCGTCTCCAGCATGAGCCAGCCCTGTAAACA 1876
Qy 1988 tgaatccctccagcaagtggaataattggcaagtggagtgatgtaagttagtctcacatgtg 2047
Db 1877 CAGAGCCCTGTCCCCAGTGGCATGTGGGCTTTGGGGCCCTGTCTCAGTACCTGTG 1936
Qy 2048 qggctcgccctacagacagagagctctctgcagcaacclgccttccagagagatgaatg 2107
Db 1937 GAGTTGAATTCAGACCCGAGATGTGTACTGTCTGCTGCACC-----CAG 1978
Qy 2108 aaacagtcactcctggtgatgagctgtgtgcgagccagcccaagccagcagcgtgcaagcct 2167
Db 1979 GGGAGACCCCTGCCCCCTCTCTGAGGAGTGGCGAGATGAAGAGCCCATGCTTACAGCAT 2038
Qy 2168 gtaacgcctttaattgccccccagcctggtacccctgcacagtgagccgtgtgtccagaa 2227
Db 2039 GCAATCAGTTGTGCTGCCCTCTCTGGTGGCACATTAAGAAATGGCAGCAGTGTTCAGGA 2098
Qy 2228 cgtgtgagcggggtgttcagaaacgtgaggttcttctgcaagcagcagcagcagcagcagc 2287

Db 2099 CTTGTGGGGGGAACACTCAGAACAGAGAGTACCTGTGCGCAGCTGCTAACGGATGGCA 2158
Qy 2288 gttctcctggaggttctcctgagacacttctgttcagcttcaaaacactgctcctccagcaagcat 2347
Db 2159 GCTTTTGAATCTCTCAGATGAATTTGTGCAAGAGCCCAAGGCATCGTCTCACAAGTCTCT 2218
Qy 2348 gcaagaagaatgactgtccagcagcagtggtctctctcagactggacagag 2397
Db 2219 GTGCCAGCAGACACTGTCTCCACATTTAGCTGTGGGAGACTGGTCTGAAG 2268

RESULT 11
AX224815

LOCUS AX224815 2175 bp DNA linear PAT 10-SEP-2001
DEFINITION Sequence 7 from Patent WO0161011.
ACCESSION AX224815
VERSION AX224815.1 GI:15554912
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (Bases 1 to 2175)
AUTHORS Donoho,G., Scoville,J., Turner,C.A., Friedrich,G., Zambrowicz,B.
and Sands,A.T.
TITLE Novel human thrombospondin repeat proteins and polynucleotides
encoding the same
JOURNAL Patent: WO 0161011-A 7 23-AUG-2001;
Lexicon Genetics Incorporated (US)
FEATURES
source 1..2175
/organism="Homo sapiens"
/db_xref="taxon:9606"
BASE COUNT 583 a 532 c 575 g 485 t
ORIGIN

Query Match 16.0%; Score 847.4; DB 6; Length 2175;
Best Local Similarity 65.1%; Pred. No. 8.4e-142;
Matches 1272; Conservative 0; Mismatches 666; Indels 15; Gaps 1;

Qy 60 cctgagttccaggaccgcagctccgagagagaccggagccctatggatgctgaggg 119
Db 186 CCAGACCTCAAGAAACACTCGTTCAGATGAGACAAAGATGGCAACTGGGATGCTTGGGG 245
Qy 120 cctgagtgatgaatgctcagcaccctgcgggggtggggtcctcactctctgagggcgtg 179
Db 246 CGACTGGAGTGACTGCTCCCGGAGCTGTGGGGAGGAGCATCATATCTCTCGCGAGATG 305
Qy 180 cctgagcagcaagagcgtggaagaataatccgatacagaagaacatgcagtaagtga 239
Db 306 TTTGACTGGAAGGAATTTGTGAAGGGCAGACATTCGGTACAGACATGCAGCAATCATGA 365
Qy 240 ctgcccaccagaaagcaggtgatattccgagctcagcaatgctcagctcataatgatgcaa 299
Db 366 CTGCCCTCCAGATCAGAGAATTTTCAGAGCCACAGCTGCTCAGCCTACATGATGATCCA 425
Qy 300 gcaccatggccagtttatgaatggttctctgtctcgaatgacctgacacccatgttc 359
Db 426 GTATCAGGGGCATTTACTATGAATGGCTTCCAGCATATAATGATCCTGCTGCCCGGTGTC 485
Qy 360 actcaagtgccaagccaaagaaacacccctgggtgtgttgactagcactaaggtcttaga 419
Db 486 ACTCAAGTGTATGCAACAGCAAAAACCTTGGTGGTGGAGCTGGCAGTAACTGACTTGA 545
Qy 420 tggtaacgctgtgtatcacagaactcttgatattgtcatcagtggtgttatgccaattgt 479
Db 546 TGGAACTCTGTGCAACAGGACCTCTTGGACATGTGTATCAGTGGCACTCTGTTCAGGCAGT 605
Qy 480 tggctgcgaatacgaagcgtgggaagcagcgtcaaggaagataaactgtgggtgtgcaacgg 539
Db 606 GGGCTGCGACTGGCACTGGGAAGCAATGCCAAGGAGGACAACTGGAGTCTGTGCCCG 665

QY	300	gacacatggccagttttatgaatgggtctctgtctaatagacctgacaaacccatgttc	359
Db	604	GTATCAGGGGCAATTACTATGAATGGCTTCCACGATATAATGATCTGCTGCCCGTGTC	663
QY	360	actcaagtgcacagcccaaaaggaacacccctggttgttgaaactlagcacctlaaggtcttaga	419
Db	664	ACTCAAGTGTCTATCACACAGGACAAACATTTGGTGGAGCTGGCACCTTAAGTACTGGA	723
QY	420	tggtaacgctgtgtctatacagaactcttggatgatgtgcatacagttgtttatgccaaattgt	479
Db	724	TGGAACTCGTTGCAACACGACACTCTTGGACATGTATACGTGGCACTTGTTCAGGCACT	783
QY	480	tggctgcgatccacagctgggaagcaccctcaaggaagataaacttggggctgcgaacgg	539
Db	784	GGGCTGGCATCGGCACTGGGAAGCAATGCCAAGGAGGAACTGTGGAGTCTGTGCCGG	843
QY	540	agatgggtccacactgcggctgtgcagaggcagtgataaaatcccagctctccgacaacaa	599
Db	844	CGATGGCTCCACCTGTCAGGCTTGTACGGGGACAATCAAGTACACAGTTCTCTCGTGA	903
QY	600	atcgggatgactgtggttgcaattccctcatggaagttagacatatctgccttgtcttaaa	659
Db	904	AAGAGAAGAAATGTAAATGCTTCTCTTTGGGAGTCGAAGTGTGAGAAATTACAGTGAA	963
QY	660	aggtcctgatcaacttatatctggaaacaaaacccctccaggagactaaagtgaaacag	719
Db	964	AGGACCTGCCACGCTCTTAAATCAAAAAACACTTCAAGGAAGCAAAAGGAGAACACAG	1023
QY	720	tctcagctccacaggaacttctctgtggacaattctatgttggaacttcagataattcc	779
Db	1024	CTTTAAACAGCCCCGGCGCTTTGTGTGTAAGAAACACACAGTGGAAATTTACAGAGGGCTC	1083
QY	780	agacaagagatactagagaatgctgtgacacctcacagcagatttctatgtccaagattcc	839
Db	1084	CGAGAGGCAAACTTTAAGATTCAGAGGACCTCTGATGGCTGATTCATCTTCAAGACCAG	1143
QY	840	taactcgggtccgcgtgacagttacagttcattctctatcaaccatcatccaccg	899
Db	1144	GTACACTTGACGCCAAACACAGCGCTGTTTCAGTTCTTTTACAGGCCATCAGTCATCA	1203
QY	900	atgagaggacagcgaattcttctctgtctcagcaacctgtggagaggattatcagctgac	959
Db	1204	GTGGAGACAAACTGACTCTTCCCTGTCACCTGTGACGTGTGGAGAGGGTTATCAGCTCAA	1263
QY	960	atcggctgagtgctacgactctgaggacgaaccggtgtgtctgacaaactctacta	1019
Db	1264	TTCTGCTGAATGTGTGSATAICCGCTTGAGAGGGTAGTTCTTGACCAATTATGTGCTACTA	1323
QY	1020	ttaccagagaaacatcaaccccaacccagcttcagagtgcaacttggatccttgtcc	1079
Db	1324	CTACCTTGAAATGTAAAAACCAAAACCAAACTGAAGAAATGCAGCATGGATCCCTGCC	1383
QY	1080	agccagtgcagatacaagcaatcatgcccattagaccttaacatcccccttcctc--gg	1137
Db	1384	ATCAAGTGTATGGATTTAAGAGATTAATGCCCTATGACCACTTCCAACTCTTCTCTCGAGC	1443
QY	1138	tgggagggcaaccctatggaccgctgctcctcctcgtgtgggggggcattccagagccgg	1197
Db	1444	TGGGAACATAATCCTTGGACTGCATGTTCCGTGTCCTGTGGAGAGGGATTCAGAGACGG	1503
QY	1198	gcagtttctgtgtgagagagacatccaggggcagtgtaacttcagtggaagagtggaaa	1257
Db	1504	AGCTTTGTGTGTAGAGGAATCCATGATGGAGAGATATTGCAGGTGGAAGAATGGAAG	1563
QY	1258	tgaatgaacccctcaagatgcccattgcgcagccctgcaacatttttgaactgccttaaa	1317
Db	1564	TGCATGTACGACCCAAACCAAGGTATGCAAACTTGTAACTCTGTTTGTATTGCCCAAG	1623
QY	1318	tggctggcacagagtggtctcgttgacagtgacatgtgcagaggccctcaataacctg	1377
Db	1624	TGNTTGCATGGAGTGGTCTCAGTGCACAGTGACTGTGGCCAGAGGGTTACGGTACCGG	1683

QY	1378	gtggctctctgcacgagccatcgaggaatgccacagaggctgtagcccaaaaacaaag	1437
Db	1684	gttggttctgtgtattaaacacccgcgagagcatgttggggctgcgaatccacaactgaag	1743
QY	1438	ccccacataaagaggaatgcctgtaccactccctgctataaacccaaagagaacctt	1497
Db	1744	ttacacatcacaagaaatgtgtcatctccatcccgctgtattaaaccaaaagaaaaagt	1803
QY	1498	ccagtcgagcccaagtgcctatggttcaacaagctcaagagctagaagaagagctgct	1557
Db	1804	ccagtggaagcaaaattgccttggctgaaacaagcacagaactagaagagaccgaata	1863
QY	1558	gtgtcaagagagccctcgttcacccagagagcctggctgcgcctgcacagtcacctgtggt	1617
Db	1864	gcaacagaagaaccaaactggtcatccagaacccctggctcagctgcagctaccacgtgtggg	1923
QY	1618	gtggggaccaggtgcgaatagtcagggtgcaggtgcctcctgtcttctcagtcctgtg	1677
Db	1924	ccaggtgtgcaggtccggcaggtgaagtgcctgtgtcctctcacattcacgcgagactgag	1983
QY	1678	gtgcacctgcctattgaagagtgtaagggcccaagccagcaatccccagctgcctgttat	1737
Db	1984	actgagctgctcccgaggaaagagtgtaaggccccaagctgccacc-----	2028
QY	1738	gcaggcccatgcagcgggaaattcctaagttcaaccagaagacagatgggctcttt	1797
Db	2029	gaacggccctgcctctctggaacatctgtatgtagagcccgccgctcccgagagctagacatc	2088
QY	1798	gdtggctctcaggaatttcagcagctgtatgactggagtagtgaggggttcaccaagtgc	1857
Db	2089	cctctcccttgaggacagtgagacgactttacgactgggagtagcgtgtggttaccctttgc	2148
QY	1858	tccgagtcctgtggaggaggtgtccagagagctgtggtgagctggttaacaaacagact	1917
Db	2149	acagcaacatgcttggaggcccatcagaaggccaatgacaggtgtgtttacatatccagacc	2208
QY	1918	eggagagcctgtgaggagaacctgtgcgtgaccagcccgccggccccacagctcctgaag	1977
Db	2209	cagcagacagtcataatgacagcttgtgtgatatgtgtccacogtctctccagccatgagccag	2268
QY	1978	tcttgcatttggatccctgcccagaaggtggga	2012
Db	2269	gcctgtataacacagagccctgtccccccaggagaga	2303

RESULT	13
LOCUS	AX224817
DEFINITION	Sequence 9 from Patent WO0161011.
ACCESSION	AX224817
VERSION	AX224817.1 GI:15554913
KEYWORDS	.
SOURCE	human.
ORGANISM	Homo sapiens
REFERENCE	Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo. (bases 1 to 1953)
AUTHORS	Donoho,G., Scoville,J., Turner,C.A., Friedrich,G., Zambrowicz,B. and Sands,A.T.
TITLE	Novel human thrombospondin repeat proteins and polynucleotides encoding the same
JOURNAL	Patent: WO 0161011-A 9 23-AUG-2001;
FEATURES	Lexicon Genetics Incorporated (US) Location/Qualifiers 1..1953 /organism="Homo sapiens" /db_xref="taxon:9606"
BASE COUNT	530 a 480 c 513 g 430 t
ORIGIN	

Query Match	14.8%;	Score 783.4;	DB 6;	Length 1953;
Best Local Similarity	64.9%;	Pred. No. 2.4e-130;		

BASE COUNT 810 a /db_xref="taxon:9606" 635 t
ORIGIN 713 c 754 g

Query Match 12.8%; Score 676.8; DB 6; Length 2912;
Best Local Similarity 58.0%; Pred. No. 2.9e-111;
Matches 1281; Conservative 0; Mismatches 882; Indels 45; Gaps 3;

Qy 717 cagttccagctccacaggaacttctctgttggaacaattctagtgtgacttccagaaatt 776
Db 1 CAGCTTTAACAGCCCGCGCTTTGTCGTAGAAAACACACAGTGGAAATTTAGAGGGG 60

Qy 777 tccagacaaagatagactagaaagtgctggaacctccacacagagatttcatgcaagat 836
Db 61 CTCCGAGAGGCAAACTTTTAAGATTCAGGCCCTTTGATGGCTGATTCATCTTCAACAC 120

Qy 837 tctgaactcgggctccgctgacagatcacagtcagattcatttctatcaaccatcatcca 896
Db 121 CAGGTACACTCGACGCAAAAGACAGCGGTTCAGTCTCTTTTACCAGGCCCATCAGTCA 180

Qy 897 ccgattgaggagagcggatttcttctgtctcagcaacctgtgagaggttatcagct 956
Db 181 TCAGTGGAGACAAACTGACTTCTTCCCTGCACGTGACGTGTGGAGAGGTTATCAGCT 240

Qy 957 gacatcggctgagtgctacagatctgagagcaacctgtgtgtgtgacccaatactgtca 1016
Db 241 CAATTCCTGCTGAATGTGTGGATATCCGCTTGAAGAGGGTAGTTCCTGACCATTATTGTCA 300

Qy 1017 ctattaccagagaaacatcaaaacccaaacaaagcttcaggagtgcaacttggatccttg 1076
Db 301 CTACTACCTCGAAATGTAAACCAAAACCAAACTGAAGGAATGCAGCATGGATCCCTG 360

Qy 1077 tccagcagtgacggtacagcagatcatcctttagacctatgacctatcaccatcccttctcg 1136
Db 361 CCCATCAAGTGTGATTTAAAGAGATAATGCCCTATGACCACACTTCCNACCTCTCTCTCG 420

Qy 1137 gtggagagcaccaccatggaccggtctctctctctgtgtgggggggcatccagagcgg 1196
Db 421 CTGGGAACATAAATCTTGGAGTGCATGTCGTCTCGTGTGGAGAGGGATTCAGAGAGC 480

Qy 1197 ggcaggttctctgtgagagagacacccagggcagtgctcaacttcagtggaagagtgaa 1256
Db 481 GAGCTTTGTGTGTAGAGAAATCCATGCATGGAGAGATATTGCAAGTGGAAAGATGGAA 540

Qy 1257 atgcatatcacccctaagatgcocactcgcgcagccctgcaacatttttgactgccctaa 1316
Db 541 GTGCATGTAGGCACCCAAACCCAAAGGTTATGCAAACTTGTAATCTGTTGATTGCCCAA 600

Qy 1317 atggtggaacagagtggtctccgtgcacagtgacatgtggtggcagggtccatgataccg 1376
Db 601 GTGATTTGCCATGGAGTGGTCTCAGTGCACAGTGACTTGTGGCCGAGGGTTACGGTACC 660

Qy 1377 tglgttctctgtatgcacatcaggaatgcacacagggaggtgtagcccaaaacaaa 1436
Db 661 GGTGTGCTGTGTATTAACCAACCGGAGAGCATGTTGGGGGCTGCAATCCACAACTGAA 720

Qy 1437 gccccacaataaagaggaatgcactgtaccocactccctgctataaacccaaagagaact 1496
Db 721 GTTACATCATAAAGAAATGTGTCATTCCATCCCTGTTATTAACCAAAAGAAAG 780

Qy 1497 tccagtcgaagcgaagtgcattggttcaaacagctcgaagtcgaagtcgaagagagtcgc 1556
Db 781 TCCAGTGGAGCAAAATGTGCTTGGCTTGAACCAAGCAAGCAAGCACTAGAAAGACCAAT 840

Qy 1557 tglgtcagagagccctcgcttcatccagagggcctgctgcctgcacagtcaccttgg 1616
Db 841 AGCAACAGAGAACCAACGTTCAITCCAGAACCTGTGCTGACCTGCTAGTACCAGGTGG 900

Qy 1617 tgtgggacccagtgcgaaatagtcaggtgccaggtgctcctgtcttctctcagtcggt 1676
Db 901 GCCGGGTGTGACGTCGTGAGGTGAAGTGCCTGTGCTCTCTACATTCACGAGACTGA 960

Qy 1677 ggctgacctgctattgacgagtggaagggcccaagccagcatccacagctgctctgta 1736
Db 961 GACTGAGTGCCTCGAGGAAGAGTGTGAAGGCCCCCAAGCTGCCACC ----- 1006

Qy 1737 tgcagggcccatgcagcggggaaattctctgagtttcaaccacagcagagacagatgggctctt 1796
Db 1007 -GAACGGCCCTCCTCTGGAAGCATGTGATGAGAGCCCGGCTCCCGAGAGCTAGACAT 1065

Qy 1797 tgggtgctgcaggaatttcagacgagctgtatgactggagtgatgggggttccacaagt 1856
Db 1066 CCCTCTCCTGAGGACAGTGAGAGCACTTACAGCTGGGAGTACGCTGGGCTTACCCCTTG 1125

Qy 1857 ctccgagtcctgtgagggaggtgtccagaggctgtgtgagctgtcttgaacaaacaaac 1916
Db 1126 CACACCAACATGCGTGGGAGGCCATCAAGAAGCCATAGCAGTGTCTTACATATCCAGAC 1185

Qy 1917 tcgggagcctgctgagggaacctgtcggtgacacgcgcgcgcgcgcgcgcgcgcgcgcgc 1976
Db 1186 CCAGCAGACAGTCAATGACAGCTTGTGTATATGTTCCACCGTCTCCAGCCATGAGCCA 1245

Qy 1977 gtcctgcaatttgatccctgcccagaaagtggaattggaaattggcaagtgagtcocatgtag 2036
Db 1246 GCCCTGTAAACACAGAGCCCTGTCCCCCAGGTGGCATGTGGGCTCTTGGGGCCCTGGTC 1305

Qy 2037 tctcacatgtgggttcggcctlacagaccagagacgtcttclgacgcacccctgttccag 2096
Db 1306 AGCTACCTGTGAGTGGAAATTCAGACCCGAGATGTGTACTGCTCCCTGCACC ----- 1355

Qy 2097 agagatgaatgaacacgtctcctggtgtgagctgtgtgcgcagcccaagccccagac 2156
Db 1356 -----CAGGGGAGACCCCTGCCCTCTCTGAGGAGTGGCGAGATGAAAGCCCCATGC 1407

Qy 2157 ggtcaagctgttaacgccttaattgccccccagcctggtaacctgcacagtggcagcc 2216
Db 1408 TTTACAGCATGCAATCAGTGTGACTGCCCTCTCTGCTGGCACATTTGAAGATGSCAGCA 1467

Qy 2217 gtgttccagaacgtgtggcgggggtgttcaagaacgtgagttcttcttcaacagcagcgt 2276
Db 1468 CTGTTCCAGGACTTGTGGCGGGGNACTCAACACAGAGATCACCTGTGGCAGCTGCT 1527

Qy 2277 ggtgatggcagctcctgagcttctctgagccttctgttcagcttcaaacctgcctg 2336
Db 1528 AACGGATGGCAGCTTTTGAATCTCAGATGAATTTGTCGAAGGACCAAGGCATCTGTC 1587

Qy 2337 ccagcaagcatgcaagaagatgactgtcccagagtggtgtctctcagactggacaga 2396
Db 1588 TCACAAGTCTCTGCCAGGACAGACTGTCTCCACATTTAGTGTGGGAGACTGGTCGAA 1647

Qy 2397 gtgttccacaagctgcggggaagcaccacagactcgaagcgcatttgccgaagatgct 2456
Db 1648 GTGTTCTGTCAAGTGTGGTGTGGAAATCCACAGAAAGCAGGAGGTGTCTCAAGGCTGGC 1707

Qy 2457 gaaaaccgctctcaacggtgttcaattccacctgtgcccgccttgccttctcttc 2516
Db 1708 AGCCAAAGGTCGGGCGCATCCCTCAGTGAGATGATGTGCAGGGATCTTACCAGGGTTCC 1767

Qy 2517 ctccatcagggcctgtatgctggc -----aacctgtgcaagggccggcgggcc 2564
Db 1768 TCTTGTAGATCTTGCAGATGCTGAGTGCGAGTAAATCAAAATCAGAGATGAAGACAAA 1827

Qy 2565 atccacaaagcacagccgcacatcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 2624
Db 1828 ACTTGGTGAGCAGGGTCCGCAGATCTCTCAGTGTCCAGAGACTCTACATTCAGACAAAGGA 1887

Qy 2625 gcagagaagcgtgcacttctgtgtgggggttcctgcctacactgtcccccaagcgcggt 2684
Db 1888 AGAGAAGCGTATTAACTGACCATTTGGTAGCAGAGCCTATTTTGCTGCCAACACATCCGT 1947

Qy 2685 ggtgctgcgtgc 2744
Db 1948 GATTATTAAAGTGGCAGTGGCAGCATTCAGAAATCTCTGATCCAGTCCGGAAGAGATGG 2007

Qy 2745 ccagcacctcatcagctcgc 2804


```

Db 2008 CCGTTGCTGCAGAACTCCAAAGCGCTGGCATCACCAAGTCAGGCTCACTAAAATCCA 2067
Qy 2805 ccgctcaagccctcgatcgagcgctcacactgctcagcgcccgccgggagca 2864
Db 2068 CGGCTTTGCTGCCCGCCGACATCGCGGTGTACCGGTGATTCGACAGGCTCTGCACAGGAAC 2127
Qy 2865 ctttgtatgaagctatcgaggaacacgcgaagctcggtgcccggcc 2912
Db 2128 AGTTGCTCAAGTCTATTGTACTGACACCGGCTCATCCGACGCC 2175

RESULT 15
LOCUS AX224811 1341 bp DNA linear PAT 10-SEP-2001
DEFINITION Sequence 3 from Patent WO0161011.
ACCESSION AX224811
VERSION AX224811.1 GI:15554910
KEYWORDS
SOURCE
ORGANISM Homo sapiens
Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 1341)
AUTHORS Donoho, G., Scoville, J., Turner, C.A., Friedrich, G., Zambrowicz, B.
and Sands, A.T.
TITLE Novel human thrombospondin repeat proteins and polynucleotides
encoding the same
JOURNAL Patent: WO 0161011-A 3 23-AUG-2001;
Lexicon Genetics Incorporated (US)
FEATURES
source
location/Qualifiers
1..1341
/organism="Homo sapiens"
/db_xref="taxon:9606"
BASE COUNT 357 a 315 c 349 g 320 t
ORIGIN

Query Match 10.5%; Score 554.8; DB 6; Length 1341;
Best Local Similarity 68.1%; Pred. No. 2.1e-89;
Matches 787; Conservative 0; Mismatches 367; Indels 2; Gaps 1;

Qy 60 cctgaagtcaggacccgacgctccgagggagggagccgagccgctatggatgcctggg 119
Db 186 CCAGACCTCAAGAAACATCTGTCAGATGAAGACAAAGATGGCAACTGGGATGCTGGG 245
Qy 120 cccatgagtgaaatgctcacacacccgctgggggtgggctcctactctctgagcgctg 179
Db 246 CGACTGGAGTGACTGCTCCCGACCTGGGGGAGGAGCATCATATTTCTCGGGAGATG 305
Qy 180 cctgagcagcaagagctgtggaagaagaataatccgatacagaacatgcagtaattgga 239
Db 306 TTTGACTGGAAGGAATTGTGAAGGGGACAGCAATTCGGTACAAGACATCGCAATCATGA 365
Qy 240 ctgcccaccagaagcagggtgattccgagctcagcaatgctcagctcagctcagctcag 299
Db 366 CTGCGCTCCAGATCGAAGATTTTCAGAGCCGACAGTGCCTCAGCCTACAAATGATGTCCA 425
Qy 300 gcacatggccagtttatgatgaatggcttcctgtgcttaatgacccctgacacccatgtc 359
Db 426 GTATCAGGGGATTTACTATGATGCTTCCAGGATATATATGATCTGCTGCCCGCTGTGC 485
Qy 360 actcaagtgcacaaaggaacacccctggtgttgtaactagcacctaaggtcttaga 419
Db 486 ACTCAAGTGTCTACAGACAGCAAAACTTGGTGTGGAGTGGACCTTAAGGTACTGGA 545
Qy 420 tggtagcgtgtgtatatacagaatctllggatgctcagctcagctcagctcagctcag 479
Db 546 TGGAACTCGTTTGAACACGGGACTCCTTGGACATGTGTATCATGTGGCATCTGTTCAGGCA 605
Qy 480 tggctgcacatcagctgggaagcaccgctcaaggaagataactgtgggtctgcaacgg 539
Db 606 GGGCTGCGATCGGCACATGGGANGCAATGCCAAGAGGAGACAACTGTGGAGTCTGTGCCG 565
```

```

Qy 540 agatgggtccacctgcccggctggtccggaggaggtataaataccagctctccgcaacaa 599
Db 666 CGATGGCTCCACCTGCGAGCTTGTACGGGGACAATCAAAAGTCACACGTTTCTCCCTGAAA 725
Qy 600 atcgatgatatactggttgcaattccctatggaagtagacatatctccctgtgtttaa 659
Db 726 AAGAGAAGAAAATGTAATTGCTTCTTGGGAAGTCGAAGTGTGAGAAATACAGTGA 785
Qy 660 aggtccctgacacttatctggaacacacacccctcaggggactaaagtgaacacag 719
Db 786 AGGACCTGCCACCTCTTTATTGAATCAAAAACACTTCAAGGAACAAAGGANGAACACAG 845
Qy 720 tctcagctccacagaaacttctctgttgacaattctctgttggaacttccagaaattccc 779
Db 846 CTTTAAACAGCCCGCGCTTTGTGCTAGAAAACACACAGTGGAAATTCAGAGGGGCTC 905
Qy 780 agacaagagatactgagaatggctggagaccactcacagacagatttcttcaagattgc 839
Db 906 CGAGAGGCAAACTTTTAAGATTCCAGGACCTGTGATGGCTGATTTTCATCTTCACAGCAG 965
Qy 840 taactcgggctccgctgacagtcacagtcacagtcacagtcacagtcacagtcacagtc 899
Db 966 GTACACTCGAGCCAAAGACAGCGTGTTCAGTCTCTTTTACCAGCCCATCAGTCA 1025
Qy 900 atggaggagagcggtattcttctctgtcagcaacccctgtggaggaggttatcagctgac 959
Db 1026 GTGGAGACAAACTGACTTCTTTCCCTCGCACTGTGACGTGTGGAGGAGTTCAGCTCAA 1085
Qy 960 atcggtgagtgctacgactgagggagcaacccgtgtgttctgacacatactgtcacta 1019
Db 1086 TTTCTGCTGAATGTGGGATATCCGCTTGAAGAGGGTATGTTCTGTGACCATTTATGTCAT 1145
Qy 1020 ttaccagagaaacatacaaccccaaccccaagcttcagagtgcaacttgatccctgttc 1079
Db 1146 CTACCTGAAAATGTAAACCAAAACCAAACTGAGGAAATGACGATGGATCCCTGCC 1205
Qy 1080 agccagtgacgggatacaagcagatcagctctatgacctctaccatccctctctcctc 1137
Db 1206 ATCAAGTGTATGGATTTTAAAGAGATAATGCCCTATGACCACTTCCAACCTCTTCTCGAG 1265
Qy 1138 tgggagggccacccatggagccgctgctcctcctcctcctcctcctcctcctcctcct 1197
Db 1266 TGGGAACATAATPCTTTGGACTGCTATGTTCCCTGCTGTGGAGAGGGGATTCAGAGAC 1325
Qy 1198 gcagtttctcctgtgtgg 1213
Db 1326 AGCTTTGTGTGTGTAG 1341
```

Search completed: July 24, 2002, 02:20:30
Job time: 13259 sec

Db 3856 atcaacacaggaagcctgcagtcacagtcgatatagggaagcaccatcaaaaacagtcgag 3915
Qy 3901 ggagtgaaatgtgacaatcaactccaggttgcaaggagtcgctaaagctgaagtcacttg 3960
Db 3916 ggagtgaaatgtgacaatcaactccaggttgcaaggagtcgctaaagctgaagtcacttg 3975
Qy 3961 ttacaggaataaaagcacaactggctcccccacacatctgcacgaagctccttgctctc 4020
Db 3976 ttacaggaataaaagcacaactggctcccccacacatctgcacgaagctccttgctctc 4035
Qy 4021 acaaacgtctcctcctgcagtcagggcctgtactcctgcagggcgccatcttcacatgga 4080
Db 4036 acaaacgtctcctcctgcagtcagggcctgtactcctgcagggcgccatcttcacatgga 4095
Qy 4081 gagtgactagagacacccagctgctgatactagatcccccccaagtcgccacacagctg 4140
Db 4096 gagtgactagagacacccagctgctgatactagatcccccccaagtcgccacacagctg 4155
Qy 4141 gaagacatcaggcctgctgcctgcacactggaccgaactctcctcagtgctgactct 4200
Db 4156 gaagacatcaggcctgctgcctgcacactggaccgaactctcctcagtgctgactct 4215
Qy 4201 cctctgggaacacagctggctcctggatcctggatcctggatcctggatcctggatcct 4260
Db 4216 cctctgggaacacagctggctcctggatcctggatcctggatcctggatcctggatcct 4275
Qy 4261 aaaggtcacctgtccctaatatcacctgtttcatgtgtcagccaattgtcactgccc 4320
Db 4276 aaaggtcacctgtccctaatatcacctgtttcatgtgtcagccaattgtcactgccc 4335
Qy 4321 acaggaactgaagcgcacatcttgagctggacagactccttcagttgcaaaccttagc 4380
Db 4336 acaggaactgaagcgcacatcttgagctggacagactccttcagttgcaaaccttagc 4395
Qy 4381 ggtgggtcgaagggaatcaagctgctcagtcagatcagatcagatcagatcagatcag 4440
Db 4396 ggtgggtcgaagggaatcaagctgctcagtcagatcagatcagatcagatcagatcag 4455
Qy 4441 aagcatcttagtgcacagactcagctggtgctggtgctggtgctggtgctggtgctg 4500
Db 4456 aagcatcttagtgcacagactcagctggtgctggtgctggtgctggtgctggtgctg 4515
Qy 4501 gctcctgtgttaacccgggggttccagcagcccgcttgagtgctgctgctgctgctgctg 4560
Db 4516 gctcctgtgttaacccgggggttccagcagcccgcttgagtgctgctgctgctgctg 4575
Qy 4561 gaggtcaacctgcccactgcgaggaaggttcgcccctgcggtgcagcccatcgctgc 4620
Db 4576 gaggtcaacctgcccactgcgaggaaggttcgcccctgcggtgcagcccatcgctgc 4635
Qy 4621 aacggagagactgccctctcgttgatggtgagctcctgctgctgctgctgctgctgctg 4680
Db 4636 aacggagagactgccctctcgttgatggtgagctcctgctgctgctgctgctgctgctg 4695
Qy 4681 tgtgggggaggtgtccagaccgaggggtgacactgtcaaaagctgaaagcctctctggatc 4740
Db 4696 tgtgggggaggtgtccagaccgaggggtgacactgtcaaaagctgaaagcctctctggatc 4755
Qy 4741 tccacctctgttccaatgacatgtgacccaggggtcgccagcggcctgtgacacccag 4800
Db 4756 tccacctctgttccaatgacatgtgacccaggggtcgccagcggcctgtgacacccag 4815
Qy 4801 gctgttaaccagagctgtgtgtgagtggtggcctctcagctggggccagtgcaatggg 4860
Db 4816 gctgttaaccagagctgtgtgtgagtggtggcctctcagctggggccagtgcaatggg 4875
Qy 4861 ccttgatcgggctcactcagctgtgacacacagacagactctctgcagacacagggat 4920
Db 4876 ccttgatcgggctcactcagctgtgacacacagacagactctctgcagacacagggat 4935
Qy 4921 ggcatacacttaccatcagagagctgactctcctccagggcctgtgagcaccacagac 4980
Db 4936 ggcatacacttaccatcagagagctgactctcctccagggcctgtgagcaccacagac 4995

Qy 4981 tgcgttgacagggcctgcagtgctacactggagagtcagggcctgtggaccctgtgcacagct 5040
Db 4996 tgcgttgacagggcctgcagtgctacactggagagtcagggcctgtggaccctgtgcacagct 5055
Qy 5041 acctgtggcaactacggcttcacagtcacggcgtgtgagtgctgcatgcccgacacac 5100
Db 5056 acctgtggcaactacggcttcacagtcacggcgtgtgagtgctgcatgcccgacacac 5115
Qy 5101 aagcagtgctgcagcactgtgctcctggggggcccgccgctgccaactggcagcgtgc 5160
Db 5116 aagcagtgctgcagcactgtgctcctggggggcccgccgctgccaactggcagcgtgc 5175
Qy 5161 aacatcaccccatgtgaaacatggagtgagagacacacacacaggtactgcgagaagtg 5220
Db 5176 aacatcaccccatgtgaaacatggagtgagagacacacacacaggtactgcgagaagtg 5235
Qy 5221 aacagctgaaactcctgcaactcagcagtttaactcctgctgtggaacttggtgc 5280
Db 5236 aacagctgaaactcctgcaactcagcagtttaactcctgctgtggaacttggtgc 5295
Qy 5281 aaagcgtga 5289
Db 5296 aaagcgtga 5304

RESULT 2
IDA06468
AB 06468 standard; cDNA; 5769 BP.
XX ABA06468;
AC ABA06468;
XX 10-JAN-2002 (first entry)
XX Human cDNA SEQ ID NO: 134.
XX Human; gene therapy; neural disorder; immune system disorder;
XX muscular disorder; reproductive disorder; gastrointestinal disorder;
XX pulmonary disorder; cardiovascular disorder; renal disorder;
XX proliferative disorder; inflammation; ss.
XX Homo sapiens.
XX WO200154474-A2.
XX PD 02-AUG-2001.
XX 17-JAN-2001; 2001WO-US01349.
XX 31-JAN-2000; 2000US-179065P.
PR 04-FEB-2000; 2000US-180628P.
PR 24-FEB-2000; 2000US-184664P.
PR 02-MAR-2000; 2000US-186350P.
PR 16-MAR-2000; 2000US-189874P.
PR 17-MAR-2000; 2000US-190076P.
PR 18-MAR-2000; 2000US-198123P.
PR 19-MAY-2000; 2000US-205515P.
PR 07-JUN-2000; 2000US-209467P.
PR 28-JUN-2000; 2000US-214886P.
PR 30-JUN-2000; 2000US-215135P.
PR 07-JUL-2000; 2000US-216647P.
PR 07-JUL-2000; 2000US-216880P.
PR 11-JUL-2000; 2000US-217487P.
PR 11-JUL-2000; 2000US-217496P.
PR 14-JUL-2000; 2000US-218290P.
PR 26-JUL-2000; 2000US-220963P.
PR 26-JUL-2000; 2000US-220964P.
PR 14-AUG-2000; 2000US-224518P.
PR 14-AUG-2000; 2000US-224519P.
PR 14-AUG-2000; 2000US-225213P.
PR 14-AUG-2000; 2000US-225214P.
PR 14-AUG-2000; 2000US-225266P.
PR 14-AUG-2000; 2000US-225267P.

124 HP-Abolic
C. elegans
protein

PR	14-AUG-2000;	2000US-22526880;	2000US-22526880P
PR	14-AUG-2000;	2000US-22527070P	2000US-22527070P
PR	14-AUG-2000;	2000US-22544470P	2000US-22544470P
PR	14-AUG-2000;	2000US-22557570P	2000US-22557570P
PR	14-AUG-2000;	2000US-22557580P	2000US-22557580P
PR	14-AUG-2000;	2000US-22557590P	2000US-22557590P
PR	18-AUG-2000;	2000US-22667610P	2000US-22667610P
PR	22-AUG-2000;	2000US-2266811P	2000US-2266811P
PR	22-AUG-2000;	2000US-22686860P	2000US-22686860P
PR	22-AUG-2000;	2000US-22711820P	2000US-22711820P
PR	30-AUG-2000;	2000US-22700090P	2000US-22700090P
PR	30-AUG-2000;	2000US-22892820P	2000US-22892820P
PR	01-SEP-2000;	2000US-22929870P	2000US-22929870P
PR	01-SEP-2000;	2000US-2293430P	2000US-2293430P
PR	01-SEP-2000;	2000US-2293444P	2000US-2293444P
PR	01-SEP-2000;	2000US-2293450P	2000US-2293450P
PR	05-SEP-2000;	2000US-2295090P	2000US-2295090P
PR	05-SEP-2000;	2000US-22951330P	2000US-22951330P
PR	06-SEP-2000;	2000US-23043730P	2000US-23043730P
PR	08-SEP-2000;	2000US-23043880P	2000US-23043880P
PR	08-SEP-2000;	2000US-23112420P	2000US-23112420P
PR	08-SEP-2000;	2000US-23125430P	2000US-23125430P
PR	08-SEP-2000;	2000US-23124440P	2000US-23124440P
PR	08-SEP-2000;	2000US-2314130P	2000US-2314130P
PR	08-SEP-2000;	2000US-23141410P	2000US-23141410P
PR	08-SEP-2000;	2000US-2320800P	2000US-2320800P
PR	12-SEP-2000;	2000US-2320810P	2000US-2320810P
PR	14-SEP-2000;	2000US-23319680P	2000US-23319680P
PR	14-SEP-2000;	2000US-23329070P	2000US-23329070P
PR	14-SEP-2000;	2000US-23339890P	2000US-23339890P
PR	14-SEP-2000;	2000US-23339990P	2000US-23339990P
PR	14-SEP-2000;	2000US-2334200P	2000US-2334200P
PR	14-SEP-2000;	2000US-2334210P	2000US-2334210P
PR	14-SEP-2000;	2000US-2330630P	2000US-2330630P
PR	14-SEP-2000;	2000US-2330640P	2000US-2330640P
PR	21-SEP-2000;	2000US-23340630P	2000US-23340630P
PR	21-SEP-2000;	2000US-2342230P	2000US-2342230P
PR	21-SEP-2000;	2000US-23429740P	2000US-23429740P
PR	25-SEP-2000;	2000US-2349970P	2000US-2349970P
PR	25-SEP-2000;	2000US-2349970P	2000US-2349970P
PR	26-SEP-2000;	2000US-2349800P	2000US-2349800P
PR	27-SEP-2000;	2000US-2354840P	2000US-2354840P
PR	27-SEP-2000;	2000US-23568340P	2000US-23568340P
PR	27-SEP-2000;	2000US-23568360P	2000US-23568360P
PR	29-SEP-2000;	2000US-2363270P	2000US-2363270P
PR	29-SEP-2000;	2000US-2363670P	2000US-2363670P
PR	29-SEP-2000;	2000US-2363680P	2000US-2363680P
PR	29-SEP-2000;	2000US-2363690P	2000US-2363690P
PR	29-SEP-2000;	2000US-2368020P	2000US-2368020P
PR	02-OCT-2000;	2000US-2368030P	2000US-2368030P
PR	02-OCT-2000;	2000US-2370370P	2000US-2370370P
PR	02-OCT-2000;	2000US-2370380P	2000US-2370380P
PR	02-OCT-2000;	2000US-2370390P	2000US-2370390P
PR	13-OCT-2000;	2000US-2370400P	2000US-2370400P
PR	13-OCT-2000;	2000US-23993450P	2000US-23993450P
PR	13-OCT-2000;	2000US-2399370P	2000US-2399370P
PR	20-OCT-2000;	2000US-2409600P	2000US-2409600P
PR	20-OCT-2000;	2000US-2412210P	2000US-2412210P
PR	20-OCT-2000;	2000US-2417850P	2000US-2417850P
PR	20-OCT-2000;	2000US-2417860P	2000US-2417860P
PR	20-OCT-2000;	2000US-2417870P	2000US-2417870P
PR	20-OCT-2000;	2000US-2418000P	2000US-2418000P
PR	20-OCT-2000;	2000US-2418090P	2000US-2418090P
PR	20-OCT-2000;	2000US-2418260P	2000US-2418260P
PR	01-NOV-2000;	2000US-2446170P	2000US-2446170P
PR			

PR	08-NOV-2000;	2000US-2445277P
PR	08-NOV-2000;	2000US-2445282P
PR	08-NOV-2000;	2000US-246532P
PR	08-NOV-2000;	2000US-246609P
PR	08-NOV-2000;	2000US-246610P
PR	08-NOV-2000;	2000US-246611P
PR	08-NOV-2000;	2000US-246613P
PR	17-NOV-2000;	2000US-249207P
PR	17-NOV-2000;	2000US-249208P
PR	17-NOV-2000;	2000US-249209P
PR	17-NOV-2000;	2000US-249210P
PR	17-NOV-2000;	2000US-249211P
PR	17-NOV-2000;	2000US-249212P
PR	17-NOV-2000;	2000US-249213P
PR	17-NOV-2000;	2000US-249214P
PR	17-NOV-2000;	2000US-249215P
PR	17-NOV-2000;	2000US-249216P
PR	17-NOV-2000;	2000US-249217P
PR	17-NOV-2000;	2000US-249218P
PR	17-NOV-2000;	2000US-249244P
PR	17-NOV-2000;	2000US-249245P
PR	17-NOV-2000;	2000US-249264P
PR	17-NOV-2000;	2000US-249265P
PR	17-NOV-2000;	2000US-249297P
PR	17-NOV-2000;	2000US-249299P
PR	17-NOV-2000;	2000US-249300P
PR	01-DEC-2000;	2000US-250160P
PR	01-DEC-2000;	2000US-250391P
PR	05-DEC-2000;	2000US-251030P
PR	05-DEC-2000;	2000US-251988P
PR	05-DEC-2000;	2000US-256719P
PR	06-DEC-2000;	2000US-256719P
PR	08-DEC-2000;	2000US-251479P
PR	08-DEC-2000;	2000US-251836P
PR	08-DEC-2000;	2000US-251868P
PR	08-DEC-2000;	2000US-251869P
PR	08-DEC-2000;	2000US-251989P
PR	08-DEC-2000;	2000US-251990P
PR	11-DEC-2000;	2000US-254907P
PR	05-JAN-2001;	2001US-259678P
XX		

Rosen CA, Barash SC, Ruben SM;

WPI; 2001-476161/51.
P-PSDB; ABB10246.

Isolated nucleic acid molecule encoding an inflammation-associated polypeptide is used in preventing, treating or ameliorating a medical condition -

Claim 1; SEQ ID NO: 134; 859pp + Sequence Listing; English.

The present invention provides human cDNAs, proteins and related genomic DNAs. These can be used in the treatment of neural, immune system, muscular, reproductive, gastrointestinal, pulmonary, cardiovascular, renal and proliferative disorders and inflammation. The present sequence is a cDNA of the invention.

Sequence 5769 BP; 1299 A; 1695 C; 1628 G; 1146 T; 1 other;

Query Match	97.8%;	Score 5171.8;	DB 22;	Length 5769;
Best Local Similarity	99.0%;	Pred. No. 0;		

Matches 5234; Conservative 1; Mismatches 3; Indels 51; Gaps 1;

QY 1 atggaatgctgcgctcgggcaactcctggcacactgctcctctttctggtttctctgctc 60

65 atggaatgctgccgctcgggcaactcctcctggtcctcttctggttctctgctc 124

Qy 61 ctgagttccaggaccgcacgcctccgaggaggaacccqccctatgqgatagcctagggc 120

[illegible]

50 ccgagccgcgcgtccgagggagggacggcctatgggatgctgggc 184

QY 121 ccattgagtgaaatgctcacgacactcggggtggtggcctcctactctctgagcgctgc 180
Db 185 ccattgagtgaaatgctcacgacactcggggtggtggcctcctactctctgagcgctgc 244
QY 181 ctgagcagcaagagctgtgaagggaagaaatccgatacagaacatgcagtaattgtgac 240
Db 245 ctgagcagcaagagctgtgaagggaagaaatccgatacagaacatgcagtaattgtgac 304
QY 241 tgcacacagaagcaggtgatttccgagctcagcaatgctcagctcataatgatgtcaag 300
Db 305 tgcacacagaagcaggtgatttccgagctcagcaatgctcagctcataatgatgtcaag 364
QY 301 caccatggcagtttatagaatgcttctctgtctataatgacctgacaacccattgttca 360
Db 365 caccatggcagtttatagaatgcttctctgtctataatgacctgacaacccattgttca 424
QY 361 ctcaagtgcacagcaagcaaacacctggtgttgtaactagcaacctaaagtcttagat 420
Db 425 ctcaagtgcacagcaagcaaacacctggtgttgtaactagcaacctaaagtcttagat 484
QY 421 ggtacgcttctatacagaatcttggataatgctacagtggtttatgccaattgtt 480
Db 485 ggtacgcttctatacagaatcttggataatgctacagtggtttatgccaattgtt 544
QY 481 ggtcgtcatcaccagctgggaagcacctgcaagggaagataactgtgggtgtctgcaacgga 540
Db 545 ggtcgtcatcaccagctgggaagcacctgcaagggaagataactgtgggtgtctgcaacgga 604
QY 541 gatggttccacctgcggctggtccgagggcagtgataataatccacgctctccgcaacccaaa 600
Db 605 gatggttccacctgcggctggtccgagggcagtgataataatccacgctctccgcaacccaaa 664
QY 601 tggatgatactgtggttgcaattccctatggaaatgacatatctgcctgtctttaaaa 660
Db 665 tggatgatactgtggttgcaattccctatggaaatgacatatctgcctgtctttaaaa 724
QY 661 ggtcctgatcacttatacttgaaacccaaaacctccaggggactaaagtgaataacagt 720
Db 725 ggtcctgatcacttatacttgaaacccaaaacctccaggggactaaagtgaataacagt 784
QY 721 ctacgtccacaggaacttctcttggaacaaattctagtgtggaattccagaaatttcca 780
Db 785 ctacgtccacaggaacttctcttggaacaaattctagtgtggaattccagaaatttcca 844
QY 781 gacaaagatactagaatgctgagacacitcacagagatttcatgtcaagattcgt 840
Db 845 gacaaagatactagaatgctgagacacitcacagagatttcatgtcaagattcgt 904
QY 841 aactcgggtccgctgcagtgacgtccagttcactcttctatacaacccatcatccaccga 900
Db 905 aactcgggtccgctgcagtgacgtccagttcactcttctatacaacccatcatccaccga 964
QY 901 tggaggagacaggaattcttctctgctcagcaacctgtggaggaggttatcagctgaca 960
Db 965 tggaggagacaggaattcttctctgctcagcaacctgtggaggaggttatcagctgaca 1024
QY 961 tcggctgaagtgtacgactgaggaacacctgtggtgctgaccaaactactgtcactat 1020
Db 1025 tcggctgaagtgtacgactgaggaacacctgtggtgctgaccaaactactgtcactat 1084
QY 1021 taccagagaacatcaaaacccaaacccagcttcaggagtgcaacttgatcttcttca 1080
Db 1085 taccagagaacatcaaaacccaaacccagcttcaggagtgcaacttgatcttcttca 1144
QY 1081 gccagtgcaggtatacagcagatcatgcttattgacctataccatcccttctcgtgtg 1140
Db 1145 gcoa-----ggtgg 1153
QY 1141 gaggccacccatgagcccgctactcctcctcgtgtgggggggacccacagcccgga 1200
Db 1154 gaggccacccatgagcccgctactcctcctcgtgtgggggggacccacagcccgga 1213

QY 1201 gttctctgtgtggaggagagacatccaggggcatgtcaattcaagtggaaagatggaaatgc 1260
Db 1214 gttctctgtgtggaggagagacatccaggggcatgtcaattcaagtggaaagatggaaatgc 1273
QY 1261 atgtacacccctaaagtgtcccatcgcgagccgtgcaacatttttgactgacctaaatgg 1320
Db 1274 atgtacacccctaaagtgtcccatcgcgagccgtgcaacatttttgactgacctaaatgg 1333
QY 1321 ctggcacagagagtggtctccgtgacacagtgcacatgtggcaggggcctcagatcccggtg 1380
Db 1334 ctggcacagagagtggtctccgtgacacagtgcacatgtggcaggggcctcagatcccggtg 1393
QY 1381 gttctctgcatogacatctgaggaatgcacacagagagctgtagccccaaaaacaagccc 1440
Db 1394 gttctctgcatogacatctgaggaatgcacacagagagctgtagccccaaaaacaagccc 1453
QY 1441 cacataaagaggaatgcacatctccactcctctataaaacccaaagagaaatcca 1500
Db 1454 cacataaagaggaatgcacatctccactcctctataaaacccaaagagaaatcca 1513
QY 1501 gtcgaggccaaagtgtccatggttcaaaacagctcaagagctagaagaaggagctgtgtg 1560
Db 1514 gtcgaggccaaagtgtccatggttcaaaacagctcaagagctagaagaaggagctgtgtg 1573
QY 1561 tcaggaggacccctgttccatccacagagggcctggtcgccctgcacagctacacctgtgtgtg 1620
Db 1574 tcaggaggacccctgttccatccacagagggcctggtcgccctgcacagctacacctgtgtgtg 1633
QY 1621 gggacccaggtgcgaatagtcagggtgcagggtgctcctgttctctcagtcctggct 1680
Db 1634 gggacccaggtgcgaatagtcagggtgcagggtgctcctgttctctcagtcctggct 1693
QY 1681 gacctgcctattgacgagtgtaaggggccaaagccagacatccacagctgctctgttatgca 1740
Db 1694 gacctgcctattgacgagtgtaaggggccaaagccagacatccacagctgctctgttatgca 1753
QY 1741 gggccatgcagcgggggaaatctcctgagttcaacccacagacagacagatgggtctttggt 1800
Db 1754 gggccatgcagcgggggaaatctcctgagttcaacccacagacagacagatgggtctttggt 1813
QY 1801 ggcctgcagattctgacagctgtatgactggaggtatgaggggttccacaaagtctcc 1860
Db 1814 ggcctgcagattctgacagctgtatgactggaggtatgaggggttccacaaagtctcc 1873
QY 1861 gagtctgtggaggaggtgtccaggaggtgtgtgagctgtgtgagctgtgaacaaacagatcgg 1920
Db 1874 gagtctgtggaggaggtgtccaggaggtgtgtgagctgtgtgagctgtgaacaaacagatcgg 1933
QY 1921 gacctgctgagagagaacctgtgctgacacagccgcggccccacagctcctgaagtcc 1980
Db 1934 gacctgctgagagagaacctgtgctgacacagccgcggccccacagctcctgaagtcc 1993
QY 1981 tgcaatttgatccctgccagcaaggtgggaaattggcaagtggagttccatgtactc 2040
Db 1994 tgcaatttgatccctgccagcaaggtgggaaattggcaagtggagttccatgtactc 2053
QY 2041 acatgtgggttcggcctacagacagagacgtcttctgacggccacctgtttccagagag 2100
Db 2054 acatgtgggttcggcctacagacagagacgtcttctgacggccacctgtttccagagag 2113
QY 2101 atgaatgaacacagatcctcctggtgatgactgtgtgcacacccaaagccagcaggtg 2160
Db 2114 atgaatgaacacagatcctcctggtgatgactgtgtgcacacccaaagccagcaggtg 2173
QY 2161 caagcttgtaacccgtttaaattgccccccagcctgtgtacccctgcacagtggagccgtgt 2220
Db 2174 caagcttgtaacccgtttaaattgccccccagcctgtgtacccctgcacagtggagccgtgt 2233
QY 2221 tccagaacgttgtggcgggggtgttccagaaacgttgaggttctttgcaagcagcagatggct 2280
Db 2234 tccagaacgttgtggcgggggtgttccagaaacgttgaggttctttgcaagcagcagatggct 2293
QY 2281 gatggcagcttctggagcttctcctgagaccttctgtttcagcttcaaaaacctgctgccag 2340

protein coordinate data; infection; ss.

Homo sapiens.
WO200121658-A1.

29-MAR-2001.

22-SEP-2000; 2000WO-US26013.

24-SEP-1999; 99US-0155709.

(HUMA-) HUMAN GENOME SCI INC.

Ni J, Baker KP, Birse CE, Ebner R, Fiscella M, Komatsoulis GA;
Lafleur DW, Moore PA, Olsen HS, Rosen CA, Ruben SA, Soppet DR;
Young PE, Wei P, Florence KA;

WPI; 2001-235311/24.

Nucleic acids encoding 32 human secreted polypeptides, useful for
preventing, diagnosing and/or treating e.g. cancers, Parkinson's
disease and diabetic retinopathy -

Claim 1; Page 747-748; 890pp; English.

The present sequence encodes one of 32 novel human secreted polypeptides.
The nucleic acid molecules and polypeptides they encode may be used in
the prevention, diagnosis and treatment of diseases such as
immune disorders (e.g. multiple sclerosis, systemic lupus erythematosus
and human immuno-deficiency virus (HIV) infections), hyperproliferative
disorders (e.g. cancers and Gaucher's disease), cardiovascular diseases
(e.g. Schmitz syndrome, Chaga's cardiomyopathy and coronary
arteriosclerosis), angiogenic disorders (e.g. corneal graft
neovascularisation and diabetic retinopathy), neurological disorders
(e.g. Huntington's chorea, Alzheimer's disease and Parkinson's disease),
infectious diseases and/or for promoting wound healing, regeneration
and/or chemotaxis. The nucleic acid molecules may be used to produce the
secreted polypeptides. They may also be used as DNA probes in diagnostic
assays to detect and quantitate the presence of similar nucleic acid
sequences in samples. The polypeptides may be used as antigens in the
production of antibodies and in assays to identify modulators of
their expression and activity.

Sequence 3559 BP; 807 A; 1037 C; 1005 G; 698 T; 12 other;

Query Match 64.1%; Score 3390.4; DB 22; Length 3559;
Best Local Similarity 97.6%; Pred. No. 0;
Matches 3510; Conservative 11; Mismatches 18; Indels 56; Gaps 6;

QY 222 aacatgcagtaattgtggaactgcccacgaagaagcaggtgatttcgaggtcagcaatgctc 281
Db 18 aacatgcagccctttggaatgcccacaaaagcmsgtgatttcgaggtcagcaatgctc 77
QY 282 agctcataatgatgtcaagcaccatgcccagttttatgaatgcttcctgtctaatga 341
Db 78 agctcataatgatgtcaagcaccatgcccagttttatgaatgcttcctgtctaatga 137
QY 342 cccatgacacccatgttcaactcaagtgcacgaagcaaacacccctgtgttggaact 401
Db 138 cccatgacacccatgttcaactcaagtgcacgaagcaaacacccctgtgttggaact 197
QY 402 agcacctaaagttcttagatggtacgcggttgcctatacagaattcttgatgtgcatcag 461
Db 198 agcacctaaagttcttagatggtacgcggttgcctatacagaattcttgatgtgcatcag 257
QY 462 tggttatgcacaaattgttgctgcgcatccagctggaagcaccctgcaagaagataa 521
Db 258 tggttatgcacaaattgttgctgcgcatccacagctggaagcaccctgcaagaagataa 317
QY 522 ctgtggggtctgcaacgagatgggtccacctgcgcgtggtccgagggcagataaatc 581

QY 4501 gctcctgtgtaaacgggggttcagcagcccgcttaagtgctgctgctgaacagcag 4560
Db 4514 gctcctgtgtaaacgggggttcagcagcccgcttaagtgctgctgctgaacagcag 4573
QY 4561 gaggtaaacctgcccactgcaggaagaggttcgcccctgcggtgcagccatcgctgc 4620
Db 4574 gaggtaaacctgcccactgcaggaagaggttcgcccctgcggtgcagccatcgctgc 4633
QY 4621 aaccgagagactgccccttctcgtggtggtgacatcctcgtgctgctgaacccgagc 4680
Db 4634 aaccgagagactgccccttctcgtggtggtgacatcctcgtgctgctgaacccgagc 4693
QY 4681 tgtggggagagtgctcagaccccgaggtgacatcctcgaagctgaaagcctctgggatac 4740
Db 4694 tgtggggagagtgctcagaccccgaggtgacatcctcgaagctgaaagcctctgggatac 4753
QY 4741 tcacccctgttccaatgacatgtgcacccaggtgcgaagcgccgtgtggacacccag 4800
Db 4754 tcacccctgttccaatgacatgtgcacccaggtgcgaagcgccgtgtggacacccag 4813
QY 4801 gctgttaaccagcagctgtgtgaggtgggcttctccagctggtggccagtgcaatggg 4860
Db 4814 gctgttaaccagcagctgtgtgaggtgggcttctccagctggtggccagtgcaatggg 4873
QY 4861 ccttgcatcgggctcactagctgtgcaacacagacagcttctccagacacgggat 4920
Db 4874 ccttgcatcgggctcactagctgtgcaacacagacagcttctccagacacgggat 4933
QY 4921 ggcataccttaccatcagagcagtgagtgcttctccagagcctgtgagcaccagaac 4980
Db 4934 ggcataccttaccatcagagcagtgagtgcttctccagagcctgtgagcaccagaac 4993
QY 4981 tgcgtggtcagagggcctgcagtgtaacactggagagtgagcctgtggacccctgtgacagct 5040
Db 4994 tgcgtggtcagagggcctgcagtgtaacactggagagtgagcctgtggacccctgtgacagct 5053
QY 5041 acctgtgcaactacgcttccagctcccgcgctgtggagtggtgcatgcccgaccacac 5100
Db 5054 acctgtgcaactacgcttccagctcccgcgctgtggagtggtgcatgcccgaccacac 5113
QY 5101 aaggcagtgctcagacacctgtgctcctgggggcccgcctgcacactgagcagcgtgc 5160
Db 5114 aaggcagtgctcagacacctgtgctcctgggggcccgcctgcacactgagcagcgtgc 5173
QY 5161 aacatcaccatgtgaaacatggagtgagagacacacacacacagctactgcgagaggtg 5220
Db 5174 aacatcaccatgtgaaacatggagtgagagacacacacacacacagctactgcgagaggtg 5233
QY 5221 aacagctgaaactgtgcaactcagccagtttaaatctcgtcgtgtggaactgtggc 5280
Db 5234 aacagctgaaactgtgcaactcagccagtttaaatctcgtcgtgtggaactgtggc 5293
QY 5281 aaagcgtga 5289
Db 5294 aaagcgtga 5302

RESULT 3
AAF97926
ID AAF97926 standard; cDNA; 3559 BP.
XX
AC AAF97926;
XX
DT 01-JUN-2001 (first entry)
XX
DE Human secreted protein cDNA, SEQ ID NO: 53.
XX
KW Human; secreted protein; immunomodulatory; antisclerotic;
KW dermatological; anti-inflammatory; anti-HIV; cytostatic; cardiant;
KW vascular; anti-angiogenic; ophthalmological; neuroprotectant;
KW nontropic; anticonvulsant; antialzheimer's; antiparkinsonian;
KW antimicrobial; vulnery; vaccine; gene therapy; cancer;

Db 318 ctgtggggtctgcaacggagatgggtccacctgcgcgctgcggagggcagtgataaatc 377
QY 582 ccagctctcgcgaacaaatcgatgatactgtgtgtgcaattccctatggaagtacaga 641
Db 378 ccagctctcgcgaacaaatcgatgatactgtgtgtgcaattccctatggaagtacaga 437
QY 642 tattcgctgtcttaaaaggtctgatcaactatatacttgaaacccaaacccctccagg 701
Db 438 tattcgctgtcttaaaaggtctgatcaactatatacttgaaacccaaacccctccagg 497
QY 702 gactaaaggtaaacagctctcagctccacaggaaacttctctgtgacaaatttagtgt 761
Db 498 gactaaaggtaaacagctctcagctccacaggaaacttctctgtgacaaatttagtgt 557
QY 762 ggaactccagaataattccagacaaagagatactgagaaatggctggaccactccacagaga 821
Db 558 ggaactccagaataattccagacaaagagatactgagaaatggctggaccactccacagaga 617
QY 822 tttaattgtcaagattcgttaactcgggctccgctgacagtacagttcaagtcttctcta 881
Db 618 tttaattgtcaagattcgttaactcgggctccgctgacagtacagttcaagtcttctcta 677
QY 882 tcaacctatccacogattggaggagacggattttcttctctgctcagcaacctgtgg 941
Db 678 tcaacctatccacogattggaggagacggattttcttctctgctcagcaacctgtgg 737
QY 942 agagattatcagctgacatcgcgtgagtgtaogatactgagagcaacctgtgtgttc 1001
Db 738 agagattatcagctgacatcgcgtgagtgtaogatactgagagcaacctgtgtgttc 797
QY 1002 tgaccaatactgtcaactattaccaagaagaacatacaaaccaaaccaagttcaggagtg 1061
Db 798 tgaccaatactgtcaactattaccaagaagaacatacaaaccaaaccaagttcaggagtg 857
QY 1062 caacttggatcctgttccagccagtgacggatatacaagcagatcatgcttatgacctta 1121
Db 858 caacttggatcctgttccagcca----- 880
QY 1122 ccataccctctcctgtgtggaggccaccctccatggacoggtgtctctcctcgtgtgggg 1181
Db 881 -----ggtggaggccaccctccatggacoggtgtctctcctcgtgtggggg 926
QY 1182 gggcatccagagcgggagtttctctgtgtgagaggagacataccaggggcatacttc 1241
Db 927 gggcatccagagcgggagtttctctgtgtgagaggagacataccaggggcatacttc 986
QY 1242 agtggagagtggaatgcatgtacacccctaaagtgcacatccgcagccctgcaacat 1301
Db 987 agtggagagtggaatgcatgtacacccctaaagtgcacatccgcagccctgcaacat 1046
QY 1302 ttttgaactgcttaaatggctggacagagtggtctccgtgcacagtgaacatgtggcca 1361
Db 1047 ttttgaactgcttaaatggctggacagagtggtctccgtgcacagtgaacatgtggcca 1106
QY 1362 gggcctcagataccctgtgtctctcgtacatcgacatcgagaatgcacacaggagctg 1421
Db 1107 gggcctcagataccctgtgtctctcgtacatcgacatcgagaatgcacacaggagctg 1166
QY 1422 tagcccaaaaacaaagcccaataaaagagaatgcatcgtaaccactccctgctataa 1481
Db 1167 tagcccaaaaacaaagcccaataaaagagaatgcatcgtaaccactccctgctataa 1226
QY 1482 acccaagaagaacttccagtcagggccaagtggccatgggttcaaacagctcaagagct 1541
Db 1227 acccaagaagaacttccagtcagggccaagtggccatgggttcaaacagctcaagagct 1286
QY 1542 agaagaagagagctgtgtgtcagagagacccctcgttcaatccagaggcctgtgcgctg 1601
Db 1287 agaagaagagagctgtgtgtcagagagacccctcgttcaatccagaggcctgtgcgctg 1346
QY 1602 cacagtcacctgtgtgtgggagaccaggtgcgaatagtcagggtgccaggtgctcctgtc 1661
Db 1347 cacagtcacctgtgtgtgggagaccaggtgcgaatagtcagggtgccaggtgctcctgtc 1406

QY 1662 tttctcagttccgtgtgtgacctgctctattgacagagtgtaagggcccaagccagcatc 1721
Db 1407 tttctcagttccgtgtgtgacctgctctattgacagagtgtaagggcccaagccagcatc 1466
QY 1722 ccagcgtgctctgttatgcaggcccatgcagcggggaataattcctgaattcaacccaagaga 1781
Db 1467 ccagcgtgctctgttatgcaggcccatgcagcggggaataattcctgaattcaacccaagaga 1526
QY 1782 gacagatgggtctcttgggtgctgcaggaatttcogacagagctgtatgactgggagatga 1841
Db 1527 gacagatgggtctcttgggtgctgcaggaatttcogacagagctgtatgactgggagatga 1586
QY 1842 ggggttcaccaagtgctccagagctcctgtgaggaagtgctccagaggcgtgtgtgagctg 1901
Db 1587 ggggttcaccaagtgctccagagctcctgtgaggaagtgctccagaggcgtgtgtgagctg 1646
QY 1902 cttgaacaaacagactcgggagcctgtgagagaaacctgtgcgtgacacagcccgccg 1961
Db 1647 cttgaacaaacagactcgggagcctgtgagagaaacctgtgcgtgacacagcccgccg 1706
QY 1962 cccacagctcctgaagtcctgcaatttggatccctgcccagcaaggtgggaaattggcaa 2021
Db 1707 cccacagctcctgaagtcctgcaatttggatccctgcccagcaaggtgggaaattggcaa 1766
QY 2022 gtggagtcctatgttagtctcacatgtgggtgcgcctcacagacagacgtctctctcag 2081
Db 1767 gtggagtcctatgttagtctcacatgtgggtgcgcctcacagacagacgtctctctcag 1826
QY 2082 ccaactgttctccagagagatgaatgaacagtcacatcctggctgatgagctgtgtccca 2141
Db 1827 ccaactgttctccagagagatgaatgaacagtcacatcctggctgatgagctgtgtccca 1886
QY 2142 gcccaagcccagacogtgcgaagcttgaacogctttaattgccccccagcctgtgtacc 2201
Db 1887 gcccaagcccagacogtgcgaagcttgaacogcttgaacogcttgaacogctgtgtacc 1946
QY 2202 tgcacagtgcagccgtgttccagaaacgtgtggcgggggtgttcaagaaacgtgaggtct 2261
Db 1947 tgcacagtgcagccgtgttccagaaacgtgtggcgggggtgttcaagaaacgtgaggtct 2006
QY 2262 ttgcaagcagcgtgctgctgacgtgcagccttccaggagcttccagcagtgaggtct 2321
Db 2007 ttgcaagcagcgtgctgctgacgtgcagccttccaggagcttccagcagtgaggtct 2066
QY 2322 ttcaaaacctgctgcagcaagcagtcgaagaagatgactgtccagcagtgaggtct 2381
Db 2067 ttcaaaacctgctgcagcaagcagtcgaagaagatgactgtccagcagtgaggtct 2126
QY 2382 ctgagactggacagagtggttccacaagctgcggggaaggcacccagactcgaagcgccat 2441
Db 2127 ctgagactggacagagtggttccacaagctgcggggaaggcacccagactcgaagcgccat 2186
QY 2442 ttgcccgaagaagtgcataaaacccgctctcaacggttgcataatccacccctgtgccg 2501
Db 2187 ttgcccgaagaagtgcataaaacccgctctcaacggttgcataatccacccctgtgccg 2246
QY 2502 cctgccccttcttctccatcagggccctgtatgtgtggaacctgtgcaagggccggcg 2561
Db 2247 cctgccccttcttctccatcagggccctgtatgtgtggaacctgtgcaagggccggcg 2306
QY 2562 gcatccagaaagcacagccgcacatcgccgcgcaggaaggttcatatccagactcg 2621
Db 2307 gcatccagaaagcacagccgcacatcgccgcgcaggaaggttcatatccagactcg 2366
QY 2622 caggcagaggaagctgcacttctgtgtgtggggccttccgctacactgctcccaagacgc 2681
Db 2367 caggcagaggaagctgcacttctgtgtgtggggccttccgctacactgctcccaagacgc 2426
QY 2682 ggt 2741
Db 2427 ggt 2486

QY	2742	cgccagcaccctcatcagctcagcgcagctcaagctgagcccttcggtatctcaagat	2801
Db	2487		
		cgccagcacctcatcagctcagcgcagctcacgttgcccccttcggcta tctcaagat	2546
QY	2802	ccacgcctcaagccctcgatgcaagcgctctacacctgctcagcgggcccgcccgga	2861
Db	2547		
		ccacgcctcaagccctcgatgcaagcgctctacacctgctcagcgggcccgcccgga	2606
QY	2862	gcactttgtattaaagctcaltcgagggaacccgcaagctcgttgccccgccttgagccc	2921
Db	2607		
		gcactttgtattaaagctcaltcgagggaacccgcaagctcgttgccccgccttgagccc	2666
QY	2922	gagaaagtgaagaagagtgctctgcggggagaagggcgcccgcaagaagagccctgcagac	2981
Db	2667		
		gagaaagtgaagaagagtgctctgcggggagaagggcgcccgcaagaagagccctgcagac	2726
QY	2982	ccacaaacacagaaacggagatcttctccaaaggcagcaagcgagagaagcgggcctggc	3041
Db	2727		
		ccacaaacacagaaacggagatcttctccaaaggcagcaagcgagagaagcgggcctggc	2786
QY	3042	cgccaaacccggggagccgctacgacgacctcgtctcccggctgcttgagacagggcgctg	3101
Db	2787		
		cgccaaacccggggagccgctacgacgacctcgtctcccggctgcttgagacagggcgctg	2846
QY	3102	gcccgagagctgctggcctcgtggaggcgagagctccgcggaaggaacacgacctc	3161
Db	2847		
		gcccgagagctgctggcctcgtggaggcgagagctcgtcggaaggaacacgacctc	2906
QY	3162	ggaggaggaccgggtgcagagcaagtgcctcgacctggccttcaccatggtagccga	3221
Db	2907		
		ggaggaggaccgggtgcagagcaagtgcctcgacctggccttcaccatggtagccga	2966
QY	3222	ggaggcgccctggagacacatcctggggaaacctctccagagccgagagagctgcgca	3281
Db	2967		
		ggaggcgccctggagacacatcctggggaaacctctccagagccgagagagctgcgca	3026
QY	3282	cctctacagcaagcacctggtggccagctggcccgaggagatcttcgcgagccacatgga	3341
Db	3027		
		cctctacagcaagcacctggtggccagctggcccgagagatcttcgcgagccacatgga	3086
QY	3342	gcaccaggacgcctctgaagccctcgagcgagagacttcccagtaactctctcgcc	3401
Db	3087		
		gcaccaggacgcctctgaagccctcgagcgagagacttcccagtaactctctcgcc	3146
QY	3402	tcataaacacgtgtctgctctcagcagctccccgcgagactctccacggggagcgcg	3461
Db	3147		
		tcataaacacgtgtctgctctcagcagctccccgcgagactctccacggggagcgcg	3206
QY	3462	gggagcctctgaagccacacgcgaagccacacctctctgcgaagatctcagcgccca	3521
Db	3207		
		gggagcctctgaagccacacgcgaagccacacctctctgcgaagatctcagcgccca	3266
QY	3522	gcagctctcagcctcgaagagtggtcaaccacctggggcagacagtggtgcccgtgccagcg	3581
Db	3267		
		gcagctctcagcctcgaagagtggtcaaccacctggggcagacagtggtgcccgtgccagcg	3326
QY	3582	gcacctgagtgcttctctgacctgtagggccatcgccaccccaagcgcttaccalcaagctg	3641
Db	3327		
		gcacctgagtg- tctctgacctgtgagcca tggcnaaccaaggcctaccatcagctg	3395
QY	3642	ggccggaatggaagaagtgtaagttcaagtgacagagattctctacagccagatgattc	3701
Db	3386		
		ggccggaatggaagaag- tcaagttcagtgacagagattctctacagscagatgattc	3444
QY	3702	cttacagatctt- ggccaccagtg- aagcagatgtaggtttctacacttgcaattgcacc	3759
Db	3445		
		yttacagatcttggccaccagtggaagcagatggtgggtttctacacttgcaattggca- c	3503
QY	3760	aatgcttgggatacagactctgtctccattgcctgcacatgaaggaagccac	3814
Db	3504		
		aatgcttgggatacagactgtctccattgcctgcacatgaaggaagccac	3558

RESULT	4
AAA47458	
ID	AAA47458 standard; cDNA; 2689 BP.
XX	
AC	AAA47458;
XX	
XX	20-OCT-2000 (first entry)
DT	
XX	
DE	Human TANGO 224 coding sequence.
XX	
XX	TANGO; 128; 140; 197; 212; 213; 224; 239; modulating agent; asthma;
KW	graft versus-host diseases; rheumatoid arthritis; psoriasis;
KW	inflammatory bowel disease; septic shock; ulcerative colitis;
KW	Crohn's disease; chronic myelogenous leukemia; cancer; liver
KW	disease; Hodgkin's disease; osteoarthritis; Lyme's disease;
KW	cachexia; autoimmune disease; myasthenia gravis; autoimmune diabetes;
KW	systemic lupus erythematosus; transgenic animal; diagnosis;
KW	prognosis; prophylactic; therapeutic; human; ds.
XX	
OS	Homo sapiens.
XX	
XX	
FH	Key Location/Qualifiers
FT	1..1443
FT	CD5
FT	/*tag= a
FT	/product= TANGO 224
XX	
XX	WO2000039284-A1.
PN	
XX	
PD	06-JUL-2000.
XX	
XX	
PF	23-DEC-1999; 99WO-US31025.
XX	
PR	30-DEC-1998; 98US-0223546.
XX	
PA	(MILL-) MILLENNIUM PHARM INC.
XX	
PI	Holtzman DA;
XX	
XX	
DR	WPI: 2000-465743/40.
DR	P-PSDB; AAB01425.
XX	
PT	Novel nucleic acid sequences encoding TANGO-128, 140, 197, 212, 213, 224 and 239 polypeptides useful for the treatment of asthma, rheumatoid arthritis, psoriasis and autoimmune diseases
PT	
PT	Claim 1; Fig 7; 209pp; English.
XX	
XX	Nucleic acids encoding TANGO polypeptides are useful as modulating agents for regulating cellular processes like asthma, graft
CC	versus-host diseases, rheumatoid arthritis, psoriasis, inflammatory
CC	bowel disease, septic shock, ulcerative colitis, Crohn's disease,
CC	chronic myelogenous leukemia, cancer, liver disease, Hodgkin's
CC	disease, osteoarthritis, Lyme's disease, cachexia and autoimmune
CC	diseases e.g. myasthenia gravis, autoimmune diabetes and systemic
CC	lupus erythematosus. The nucleic acids are also useful for producing
CC	transgenic animals and the TANGO polypeptides themselves. Partial
CC	TANGO-128, 140, 197, 212, 213, 224, 239 sequences are useful in
CC	forensic biology, for diagnostic assays, prognostic assays,
CC	pharmacogenomics and for monitoring clinical trials. TANGO
CC	polypeptides are suitable for both prophylactic and therapeutic
CC	methods for treating a subject at risk of a disorder or having a
CC	disorder associated with aberrant TANGO expression. A wide range
CC	of cellular disorders can be treated.
XX	
XX	Sequence 2689 BP; 633 A; 740 C; 734 G; 582 T; 0 other;
SQ	

Query Match	48.3%;	Score 2556.2;	DB 21;	Length 2689;
Best Local Similarity	98.9%;	Pred. No. 0;		
Matches 2594;	Conservative 0;	Mismatches 3;	Indels 26;	Gaps
QY	1	atggaatgctgcgcgtgcgggcgaactcctgcgaactgctcctctttctggtcttctgcgc	60	


```
Qy 2195 ggtacctgcacagtgcagccgtgttccagaacagtgtgcccgggtgttccagaacagt 2254
Db 2287 ggtacctgcacagtgcagccgtgttccagaacagtgtgcccgggtgttccagaacagt 2346
Qy 2255 aggtctttgcaagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 2314
Db 2347 aggtctttgcaagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 2406
Qy 2315 gttcagcttcaaaacctgcctgcagcagcagcagcagcagcagcagcagcagcagc 2374
Db 2407 gttcagcttcaaaacctgcctgcagcagcagcagcagcagcagcagcagcagcagc 2466
Qy 2375 ggtctctcagactgcacagagtgttccacaacagtgcgggaagcagccagactcga 2434
Db 2467 ggtctctcagactgcacagagtgttccacaacagtgcgggaagcagccagactcga 2526
Qy 2435 gcgcatttgcgaagaatgctgaacccgcctctcaacggtgttcaattccacctgt 2494
Db 2527 gcgcatttgcgaagaatgctgaacccgcctctcaacggtgttcaattccacctgt 2586
Qy 2495 gccgcacctgccttctctctccatcagccagccctgtatctggtgcaacctgtgcaagg 2554
Db 2587 gccgcacctgccttctctctccatcagccagccctgtatctggtgcaacctgtgcaagg 2646
Qy 2555 ccggggcgccatccacgaagcacagccgcacatcgccgcgc 2597
Db 2647 ccggggcgccatccacgaagcacagccgcacatcgccgcgc 2689

RESULT 5
AAA47482
ID AAA47482 standard; cDNA; 2689 BP.
AC AAA47482;
XX
DT 20-OCT-2000 (first entry)
DE Human TANGO 224 coding sequence (form 1).
XX
KW TANGO; 128; 140; 197; 212; 213; 224; 239; modulating agent; asthma;
KW graft versus-host diseases; rheumatoid arthritis; psoriasis;
KW inflammatory bowel disease; septic shock; ulcerative colitis;
KW Crohn's disease; chronic myelogenous leukemia; cancer; liver
KW disease; Hodgkin's disease; osteoarthritis; Lyme's disease;
KW cachexia; autoimmune disease; myasthenia gravis; autoimmune diabetes;
KW systemic lupus erythematosus; transgenic animal; diagnosis;
KW prognosis; prophylactic; therapeutic; human; ds.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT CDS 67..2688
FT /*tag= a
FT /product= TANGO 224
XX
FN WO200039284-A1.
XX
PD 06-JUL-2000.
XX
PF 23-DEC-1999; 99WO-US31025.
XX
PR 30-DEC-1998; 98US-0223546.
XX
PA (MILL-) MILLENNIUM PHARM INC.
XX
PI Holtzman DA;
XX
WPI: 2000-465743/40.
DR P-PSDB; AAB01431.
XX
XX Novel nucleic acid sequences encoding TANGO-128, 140, 197, 212, 213,
XX 224 and 239 polypeptides useful for the treatment of asthma, rheumatoid
XX arthritis, psoriasis and autoimmune diseases
```

```
XX Claim 1; Fig 30; 209pp; English.
XX
CC Nucleic acids encoding TANGO polypeptides are useful as modulating
CC agents for regulating cellular processes like asthma, graft
CC versus-host diseases, rheumatoid arthritis, psoriasis, inflammatory
CC bowel disease, septic shock, ulcerative colitis, Crohn's disease,
CC chronic myelogenous leukemia, cancer, liver disease, Hodgkin's
CC disease, osteoarthritis, Lyme's disease, cachexia and autoimmune
CC diseases e.g. myasthenia gravis, autoimmune diabetes and systemic
CC lupus erythematosus. The nucleic acids are also useful for producing
CC transgenic animals and the TANGO polypeptides themselves. Partial
CC TANGO-128, 140, 197, 212, 213, 224, 239 sequences are useful in
CC forensic biology, for diagnostic assays, prognostic assays,
CC pharmacogenomics and for monitoring clinical trials. TANGO
CC polypeptides are suitable for both prophylactic and therapeutic
CC methods for treating a subject at risk of a disorder or having a
CC disorder associated with aberrant TANGO expression. A wide range
CC of cellular disorders can be treated.
XX
SQ Sequence 2689 BP; 633 A; 739 C; 734 G; 583 T; 0 other;
```

```
Query Match 48.3%; Score 2552.6; DB 21; Length 2689;
Best Local Similarity 98.8%; Pred. No. 0;
Matches 2592; Conservative 0; Mismatches 4; Indels 27; Gaps 1;

Qy 1 atggaatgtcgctgcctggggaactctctgcacactgctctcttcttctggttctctgctc 60
Db 67 atggaatgtcgctgcctggggaactctctgcacactgctctcttcttctggttctctgctc 126
Qy 61 ctgaattccaggacgcagcgtccagagagaccgggagcgcctatggatgcctggggc 120
Db 127 ctgaattccaggacgcagcgtccagagagaccgggagcgcctatggatgcctggggc 186
Qy 121 ccatggagtgaatgctcagcgcctcgggggtggggtcctctctctctctctctctctctctc 180
Db 187 ccatggagtgaatgctcagcgcctcgggggtggggtcctctctctctctctctctctctctc 246
Qy 181 ctgagcagcagagctgtgaaggagaataatcccatagacagacatgcagtaattggac 240
Db 247 ctgagcagcagagctgtgaaggagaataatcccatagacagacatgcagtaattggac 306
Qy 241 tgcaccacagaagcaggtgattccagctcagcagcagcagcagcagcagcagcagcagc 300
Db 307 tgcaccacagaagcaggtgattccagctcagcagcagcagcagcagcagcagcagcagc 366
Qy 301 caccatggccagtttatgaatggcttctgtgtctaatgacctgacacacccatgttca 360
Db 367 caccatggccagtttatgaatggcttctgtgtctaatgacctgacacacccatgttca 426
Qy 361 ctcaagtgccaaagccaaaggaaacacccctgggtgttgtaactagcacctaaagctttagat 420
Db 427 ctcaagtgccaaagccaaaggaaacacccctgggtgttgtaactagcacctaaagctttagat 486
Qy 421 ggtacgcttgctatacagaatcttggatagtgtcagcagtggttattgccaattgtt 480
Db 487 ggtacgcttgctatacagaatcttggatagtgtcagcagtggttattgccaattgtt 546
Qy 481 ggctcgatcaccagctgggaagcagcgtcaaggaagataaactgtgggtctgcaacgga 540
Db 547 ggctcgatcaccagctgggaagcagcgtcaaggaagataaactgtgggtctgcaacgga 606
Qy 541 gatgggtccacctgcggtgttccaggaggcagtaataatccagctctctccgcaacccaaa 600
Db 607 gatgggtccacctgcggtgttccaggaggcagtaataatccagctctctccgcaacccaaa 666
Qy 601 tcgagatgactgtgttgcaattccctatggaatagacatatctgcgctgttcttaaaa 660
Db 667 tcgagatgactgtgttgcaattccctatggaatagacatatctgcgctgttcttaaaa 726
Qy 661 ggtctctgacattatctggaacccaaacccctccaggggagctaaaggtgaaacagct 720
Db 726 ggtctctgacattatctggaacccaaacccctccaggggagctaaaggtgaaacagct 780
```


XX KX cirrhosis; psoriasis; acquired immune deficiency syndrome; anaemia;
KW asthma; Crohn's disease; infection; Alzheimer's disease; schizophrenia;
KW Parkinson's disease; Huntington's diseases; ovulatory defect;
KW muscular dystrophy; ss.
XX Homo sapiens.
XX W0200000610-A2.
XX 06-JAN-2000.
XX 25-JUN-1999; 99WO-US14484.
XX 26-JUN-1998; 98US-0090762.
XX 31-JUL-1998; 98US-0054983.
XX 01-OCT-1998; 98US-0102686.
XX 11-DEC-1998; 98US-0112129.
XX (INCY-) INCYTE PHARM INC.
XX Lal P, Tang YT, Gorgone GA, Corley NC, Guegler KU, Baughn MR;
XX Akerblom IE, Au-Young J, Yue H, Patterson C, Reddy R, Hillman JL;
XX Bandman O;
XX WPI; 2000-160673/14.
XX P-PSDB; AAY87347.
XX New human signal peptide-containing proteins useful in treatment,
XX prevention and diagnosis of e.g. cancer, inflammation and
XX cardiovascular disease -
XX Claim 9; Page 321-322; 327pp; English.
XX AAY98109 to AAY98242 encode AAY87224 to AAY87357 which represent the
XX human signal peptide-containing proteins HSP-1 to HSP-134. HSPs have
XX anticancer, anti-inflammatory, antimicrobial, nootropic, hepatotropic,
XX neuroprotective, cardiovascular and antiasthmatic activities, and can
XX be used in gene therapy. HSPs can be used to treat or prevent disorders
XX associated with decreased activity or function of HSP. Antagonists of
XX HSP are used to treat or prevent disorders associated with increased
XX activity or function of HSP. Such disorders include cell proliferation
XX (including cancer), inflammation, cardiovascular, neurological,
XX reproductive or developmental disorders, (e.g. arteriosclerosis,
XX cirrhosis, psoriasis, acquired immune deficiency syndrome, anaemia,
XX asthma, Crohn's disease, microbial or other infections, congestive or
XX ischaemic heart disease, Alzheimer's, Parkinson's or Huntington's
XX diseases, schizophrenia, ovulatory defects, muscular dystrophy). HSP
XX nucleic acids can be used for the recombinant production of HSP, for
XX detecting HSP in standard hybridisation and amplification assays (for
XX diagnosis and monitoring), in gene therapy, as antisense,
XX triplex-forming or ribozyme therapeutics, for detecting related sequences
XX or genetic variations, and for chromosomal mapping. HSP are also used to
XX raise specific antibodies (Ab) and to screen for agonists and
XX antagonists (potential therapeutic agents). Ab are used to diagnose, or
XX monitor, HSP-related diseases (in usual immunoassays), as therapeutic
XX antagonists, in competitive drug screens, and for purification of HSP
XX from natural sources.
XX Sequence 3045 BP; 763 A; 772 C; 768 G; 723 T; 19 other;

QY	121	ccatggagtgatgctcagccacctcaggggtgagggcctcctactctctgagcgctgc	180
DB	250	ccatggagtgatgctcagccacctcaggggtgagggcctcctactctctgagcgctgc	309
QY	181	ctgagcagcagagctggaaggaagaataccgatacagaacaatgagtaattgtgac	240
DB	310	ctgagcagcagagctggaaggaagaataccgatacagaacaatgagtaattgtgac	369
QY	241	tgccaccagcagcaggtgattccgagctcagcaatgctcagctcataatgagtcag	300
DB	370	tgccaccagcagcaggtgattccgagctcagcaatgctcagctcataatgagtcag	429
QY	301	caccatggccagtttatgaatggcttcctgctcgaatgacccctgacaacccattgtca	360
DB	430	caccatggccagtttatgaatggcttcctgctcgaatgacccctgacaacccattgtca	489
QY	361	ctcaagtgcacagcacaaggaacacccctggttgtaactagcaccctaaagctcttagat	420
DB	490	ctcaagtgcacagcacaaggaacacccctggttgtaactagcaccctaaagctcttagat	549
QY	421	ggtacgcgttgctatcacagaatctttggatagtgtgcacagtggtgttatgcaaatgtt	480
DB	550	ggtacgcgttgctatcacagaatctttggatagtgtgcacagtggtgttatgcaaatgtt	609
QY	481	ggctgcgataccagctgggaagcaccgtcaaggaagataaactgtggggtctgcaacgga	540
DB	610	ggctgcgataccagctgggaagcaccgtcaaggaagataaactgtggggtctgcaacgga	669
QY	541	gatgggtccacctgccgcgtggtccgagggcagataaaatccagctctccgacaacaa	600
DB	670	gatgggtccacctgccgcgtggtccgagggcagataaaatccagctctccgacaacaa	729
QY	601	tcggatgatactgtggtgtaattccctatggaagtgcacatatctgcgcttcttaaaa	660
DB	730	tcggatgatactgtggtgtaattccctatggaagtgcacatatctgcgcttcttaaaa	789
QY	661	ggtcctgatacttatctgaaacaaacccctccagggagactaaaggtagaaacagt	720
DB	790	ggtcctgatacttatctgaaacaaacccctccagggagactaaaggtagaaacagt	849
QY	721	ctcagctccacaggaacttctctgtggacaattctagtgtgactccagaaattcca	780
DB	850	ctcagctccacaggaacttctctgtggacaattctagtgtgactccagaaattcca	909
QY	781	gacaaagatactgagaatggtggacacactcagcagagatttcattgcagattcgt	840
DB	910	gacaaagatactgagaatggtggacacactcagcagagatttcattgcagattcgt	969
QY	841	aactcgggtccgctcagatcagatccagttctctctatcaacccatccaccga	900
DB	970	aactcgggtccgctcagatcagatccagttctctctatcaacccatccaccga	1029
QY	901	tgagggagcagcagatttcttctcctcagcaacctgtggaggaggtatcagctgaca	960
DB	1030	tgagggagcagcagatttcttctcctcagcaacctgtggaggaggtatcagctgaca	1089
QY	961	tcggctgagtgctacgatactgagagcaacccgtgtggtgtgactgacaaactgacctat	1020
DB	1090	tcggctgagtgctacgatactgagagcaacccgtgtggtgtgactgacaaactgacctat	1149
QY	1021	taccagagacatcaaaccccaacccagcttcagagtgcaacttggaatcctgtcca	1080
DB	1150	taccagagacatcaaaccccaacccagcttcagagtgcaacttggaatcctgtcca	1209
QY	1081	gccagtgcaggtatacagcagatcattgccttatgacctaccatccctctcctcgtg	1140
DB	1210	gccagtgcaggtatacagcagatcattgccttatgacctaccatccctctcctcgtg	1269
QY	1141	gagggcaccacatgacccggtgctcctcctggtgggggggggagccagcgagca	1200
DB	1270	gagggcaccacatgacccggtgctcctcctggtgggggggggagccagcgagca	1329
QY	1201	gtttcctgtgtggaggaggacatccaggggcatgtcacttcagtgggaagagtgaatgc	1260

```

Dbb 1330 gttctctgtgtgagagagacatccaggggcatgtcacttcagtggagagtggaatgc 1389
QY 1261 atgtacacccttaagatgcccacgcgcagccctgcacacatttttgaactgccttaaatgg 1320
Dbb 1390 atgtacacccttaagatgcccacgcgcagccctgcacacatttttgaactgccttaaatgg 1449
QY 1321 ctggcacagagtggtctccgtgcacagtgacatgtggccagggcctcagataccgtgtg 1380
Dbb 1450 ctggcacagagtggtctccgtgcacagtgacatgtggccagggcctcagataccgtgtg 1509
QY 1381 gtccctctgcatcgacatcgagaaatgcacagagagtggtgtagcccaaaaaacaaagccc 1440
Dbb 1510 gtccctctgcatcgacatcgagaaatgcacagagagtggtgtagcccaaaaaacaaagccc 1569
QY 1441 cacataaagaggaatgcctgtaccaccctcctgtataaaaccccaaaagagaactcca 1500
Dbb 1570 cacataaagaggaatgcctgtaccaccctcctgtataaaaccccaaaagagaactcca 1629
QY 1501 gtccagggccaagtcccatggttcaaaacaaagctcaagagctagaagaagagctgctgtg 1560
Dbb 1630 gtccagggccaagtcccatggttcaaaacaaagctcaagagctagaagaagagctgctgtg 1689
QY 1561 tcagaggagccctcgttcaatccagagccctggtcggcctgcacagtgacactgtggtgtg 1620
Dbb 1690 tcagaggagccctcgttcaatccagagccctggtcggcctgcacagtgacactgtggtgtg 1749
QY 1621 gggaccacaggtgcgaatagtcaggtgcaggtgcctcctgctcttctcagtcagtcggct 1680
Dbb 1750 gggaccacaggtgcgaatagtcaggtgcaggtgcctcctgctcttctcagtcagtcggct 1809
QY 1681 gacctgctctattgacagtgtagaaggcccaagccagcagcagcagcagcagcagcagcagc 1740
Dbb 1810 gacctgctctattgacagtgtagaaggcccaagccagcagcagcagcagcagcagcagcagc 1869
QY 1741 gggccatcagcggggaaattcctgagttcaaccagagcagcagcagcagcagcagcagcagc 1860
Dbb 1870 gggccatcagcggggaaattcctgagttcaaccagagcagcagcagcagcagcagcagcagc 1929
QY 1801 ggcctgaagatttcagcagcagctgatgactggagtgagtgagggttcaccaagtgctcc 1860
Dbb 1930 ggcctgaagatttcagcagcagctgatgactggagtgagtgagggttcaccaagtgctcc 1989
QY 1861 gagtctctgtgagcaggtgtccagagagctgtgtgagctccttgaaacaaacagactcgg 1920
Dbb 1990 gagtctctgtgagcaggtgtccagagagctgtgtgagctccttgaaacaaacagactcgg 2049
QY 1921 gaggcc-tgctgagagagacacctgtcgtgacacagccgcccagcagcagcagcagcagcagc 1979
Dbb 2050 gaggcctgtcgtgagagacacctgtcgtgaccca-ccgccggccccccacagcagcagcagc 2108
QY 1980 ctgcaatttgatccctgcccagaag 2006
Dbb 2109 ctgcaatttgatccctgcccagaag 2135

RESULT 7
ABR06676
ID ABR06676 standard; cDNA; 2031 BP.
XX
AC ABR06676;
XX
DT 10-JAN-2002 (first entry)
DE Human cDNA SEQ ID NO: 342.
XX
KW Human; gene therapy; neural disorder; immune system disorder;
KW muscular disorder; reproductive disorder; gastrointestinal disorder;
KW pulmonary disorder; cardiovascular disorder; renal disorder;
KW proliferative disorder; inflammation; ss.
XX
OS Homo sapiens.
XX
```

```

PN
XX WO200154474-A2.
PD
XX 02-AUG-2001.
PF
XX 17-JAN-2001; 2001WO-US01349.
XX
PR 31-JAN-2000; 2000US-179065P.
PR 04-FEB-2000; 2000US-180628P.
PR 24-FEB-2000; 2000US-184664P.
PR 02-MAR-2000; 2000US-186350P.
PR 16-MAR-2000; 2000US-186874P.
PR 17-MAR-2000; 2000US-190076P.
PR 18-APR-2000; 2000US-198123P.
PR 19-MAY-2000; 2000US-205515P.
PR 07-JUN-2000; 2000US-209467P.
PR 28-JUN-2000; 2000US-214886P.
PR 30-JUN-2000; 2000US-215135P.
PR 07-JUL-2000; 2000US-216647P.
PR 07-JUL-2000; 2000US-216880P.
PR 11-JUL-2000; 2000US-217487P.
PR 11-JUL-2000; 2000US-217496P.
PR 14-JUL-2000; 2000US-218290P.
PR 26-JUL-2000; 2000US-220963P.
PR 26-JUL-2000; 2000US-220964P.
PR 14-AUG-2000; 2000US-224518P.
PR 14-AUG-2000; 2000US-224519P.
PR 14-AUG-2000; 2000US-225213P.
PR 14-AUG-2000; 2000US-225214P.
PR 14-AUG-2000; 2000US-225266P.
PR 14-AUG-2000; 2000US-225267P.
PR 14-AUG-2000; 2000US-225268P.
PR 14-AUG-2000; 2000US-225270P.
PR 14-AUG-2000; 2000US-225447P.
PR 14-AUG-2000; 2000US-225757P.
PR 14-AUG-2000; 2000US-225758P.
PR 18-AUG-2000; 2000US-225759P.
PR 18-AUG-2000; 2000US-226279P.
PR 22-AUG-2000; 2000US-226681P.
PR 22-AUG-2000; 2000US-226868P.
PR 22-AUG-2000; 2000US-227182P.
PR 23-AUG-2000; 2000US-227009P.
PR 30-AUG-2000; 2000US-228924P.
PR 01-SEP-2000; 2000US-229287P.
PR 01-SEP-2000; 2000US-229343P.
PR 01-SEP-2000; 2000US-229344P.
PR 05-SEP-2000; 2000US-229345P.
PR 05-SEP-2000; 2000US-229509P.
PR 06-SEP-2000; 2000US-229513P.
PR 06-SEP-2000; 2000US-230437P.
PR 06-SEP-2000; 2000US-230438P.
PR 08-SEP-2000; 2000US-231242P.
PR 08-SEP-2000; 2000US-231243P.
PR 08-SEP-2000; 2000US-231244P.
PR 08-SEP-2000; 2000US-231413P.
PR 08-SEP-2000; 2000US-231414P.
PR 08-SEP-2000; 2000US-232080P.
PR 08-SEP-2000; 2000US-232081P.
PR 12-SEP-2000; 2000US-231968P.
PR 14-SEP-2000; 2000US-232397P.
PR 14-SEP-2000; 2000US-232398P.
PR 14-SEP-2000; 2000US-232399P.
PR 14-SEP-2000; 2000US-232400P.
PR 14-SEP-2000; 2000US-232401P.
PR 14-SEP-2000; 2000US-233063P.
PR 14-SEP-2000; 2000US-233064P.
PR 14-SEP-2000; 2000US-233065P.
PR 21-SEP-2000; 2000US-234223P.
PR 21-SEP-2000; 2000US-234274P.
PR 25-SEP-2000; 2000US-234997P.
PR 25-SEP-2000; 2000US-234998P.
PR 26-SEP-2000; 2000US-235484P.
PR 27-SEP-2000; 2000US-235834P.
PR 27-SEP-2000; 2000US-235836P.
```


Db 729 ggagagctgactgagagcaccagctgctgatactctagatcccccccaagtccccacacag 788
QY 4138 ttggaagacatcaggcccttgcctgcactgacccgacccaaacctctctcaagtgcgaag 4197
Db 789 ttggaagacatcaggcccttgcctgcactgacccgacccaaacctctctcaagtgcgaag 848
QY 4198 tctcctctgggaacacagctgctcctggatcctcctggaaattctgctcctctggctgccc 4257
Db 849 tctcctctgggaacacagctgctcctggatcctcctggaaattctgctcctctggctgccc 908
QY 4258 atcaaaaggtccacctgctccataatcacctggtttcatggtggtgacgacaaattgtcaact 4317
Db 909 atcaaaaggtccacctgctccataatcacctggtttcatggtggtgacgacaaattgtcaact 968
QY 4318 gccacaggactgacgcacatcactctggcagctggacagatccttcaagtggcaaacctt 4377
Db 969 gccacaggactgacgcacatcactctggcagctggacagatccttcaagtggcaaacctt 1028
QY 4378 agcgggtggtctcaagggaattcagctccttgcctgctcagaatgagcagggtgtctcatg 4437
Db 1029 agcgggtggtctcaagggaattcagctccttgcctgctcagaatgagcagggtgtctcatg 1088
QY 4438 cagaaggcatctttagtatccaagattactggtgtgtctgtgacagactgccaacctgc 4497
Db 1089 cagaaggcatctttagtatccaagattactggtgtgtctgtgacagactgccaacctgc 1148
QY 4498 tcagcctcctgtgtaacgggggttcagcagcccgcttgagtgctcgtcgaacagc 4557
Db 1149 tcagcctcctgtgtaacgggggttcagcagcccgcttgagtgctcgtcgaacagc 1208
QY 4558 acggaggttcaacctgccactgcgcagggaaggttcgctcggctgagcccatgcgc 4617
Db 1209 acggaggttcaacctgccactgcgcagggaaggttcgctcggctgagcccatgcgc 1268
QY 4618 tgcacccggagagactgccccttctcgttgatggtgactcctcgtcctgctgtaacccgcg 4677
Db 1269 tgcacccggagagactgccccttctcgttgatggtgactcctcgtcctgctgtaacccgcg 1328
QY 4678 agctgtgggggaggtgtccagaccgcagggtgacctgtlcaaaagctgaaaagcctctggg 4737
Db 1329 agctgtgggggaggtgtccagaccgcagggtgacctgtlcaaaagctgaaaagcctctggg 1388
QY 4738 atctccacctgtgtccaatgacatgtgacccaggtgcgcaaggccctgtggacacc 4797
Db 1389 atctccacctgtgtccaatgacatgtgacccaggtgcgcaaggccctgtggacacc 1448
QY 4798 caggcctgttaaccagcagctgtgtgtggagtgggccttctcagctggggccagtgaat 4857
Db 1449 caggcctgttaaccagcagctgtgtgtggagtgggccttctcagctggggccagtgaat 1508
QY 4858 gggccttgcatcgggcctcaacctagctgtgcaacacagacaagtctctgccagacacgg 4917
Db 1509 gggccttgcatcgggcctcaacctagctgtgcaacacagacaagtctctgccagacacgg 1568
QY 4918 gatgcatcacccttaaccatcagacagcagctgactccttcgagcctgtgagcaccag 4977
Db 1569 gatgcatcacccttaaccatcagacagcagctgactccttcgagcctgtgagcaccag 1628
QY 4978 aactgctgttgaggcctgcagtgtaacctggagagtgagcctgtgacacctgtgcaca 5037
Db 1629 aactgctgttgaggcctgcagtgtaacctggagagtgagcctgtgacacctgtgcaca 1688
QY 5038 gctacacctgtggcaactacgcttccagtcgccggtgtggagtggtgcatgcccgcacc 5097
Db 1689 gctacacctgtggcaactacgcttccagtcgccggtgtggagtggtgcatgcccgcacc 1748
QY 5098 acaaggcagctgactgagcactgtcctctggggcccgccctgccaactggcagcgc 5157
Db 1749 acaaggcagctgactgagcactgtcctctggggcccgccctgccaactggcagcgc 1808
QY 5158 tgcacatcaccccatgtgaaaacatggagtgacagacacccaccaggtactgcgagaag 5217
Db 1809 tgcacatcaccccatgtgaaaacatggagtgacagacacccaccaggtactgcgagaag 1868

QY 5218 gtcaaacagctgaactctgccaactcagccagtttaaatctcgtcgtggaacttgt 5277
Db 1869 gtcaaacagctgaactctgccaactcagccagtttaaatctcgtcgtggaacttgt 1928
QY 5278 ggcaaaagcgtga 5289
Db 1929 ggcaaaagcgtga 1940
RESULT 8
AAZ34146
ID AAZ34146 standard; cDNA; 1869 BP.
XX AAZ34146;
AC AAZ34146;
XX
DT 07-DEC-1999 (first entry)
XX Human PRO1071 nucleotide sequence.
DE
DE Human PRO; EST; expressed sequence tag; PCR primer; hybridisation;
KW probe; blood coagulation disorder; cancer; cellular adhesion disorder;
KW secreted protein; transmembrane protein; ss.
XX Homo sapiens.
PN WO9946281-A2.
XX
PD 16-SEP-1999.
XX
PF 08-MAR-1999; 99WO-US05028.
XX
PR 10-MAR-1998; 98US-0077450.
PR 11-MAR-1998; 98US-0077632.
PR 11-MAR-1998; 98US-0077641.
PR 11-MAR-1998; 98US-0077649.
PR 12-MAR-1998; 98US-0077791.
PR 13-MAR-1998; 98US-0078004.
PR 17-MAR-1998; 98US-0040220.
PR 20-MAR-1998; 98US-0078886.
PR 20-MAR-1998; 98US-0078910.
PR 20-MAR-1998; 98US-0078936.
PR 20-MAR-1998; 98US-0078939.
PR 25-MAR-1998; 98US-0079294.
PR 26-MAR-1998; 98US-0079656.
PR 27-MAR-1998; 98US-0079663.
PR 27-MAR-1998; 98US-0079664.
PR 27-MAR-1998; 98US-0079689.
PR 27-MAR-1998; 98US-0079728.
PR 27-MAR-1998; 98US-0079786.
PR 30-MAR-1998; 98US-0079920.
PR 30-MAR-1998; 98US-0079923.
PR 31-MAR-1998; 98US-0080105.
PR 31-MAR-1998; 98US-0080107.
PR 31-MAR-1998; 98US-0080165.
PR 31-MAR-1998; 98US-0080194.
PR 01-APR-1998; 98US-0080327.
PR 01-APR-1998; 98US-0080328.
PR 01-APR-1998; 98US-0080333.
PR 01-APR-1998; 98US-0080334.
PR 08-APR-1998; 98US-0081049.
PR 08-APR-1998; 98US-0081070.
PR 08-APR-1998; 98US-0081071.
PR 09-APR-1998; 98US-0081195.
PR 09-APR-1998; 98US-0081203.
PR 09-APR-1998; 98US-0081229.
PR 15-APR-1998; 98US-0081817.
PR 15-APR-1998; 98US-0081838.
PR 15-APR-1998; 98US-0081952.
PR 15-APR-1998; 98US-0081955.
PR 21-APR-1998; 98US-0082568.
PR 21-APR-1998; 98US-0082569.
PR 22-APR-1998; 98US-0082700.


```

Db 1093 tcggctgagtgctacgatctgaggagcaaccgtgtgtgtgtagcacaatactgtcaact 1152
QY 1021 taccagagaacatacaaaacccaacagcttcaggagtgcaacttggaactctgtccca 1080
Db 1153 taccagagaacatacaaaacccaacagcttcaggagtgcaacttggaactctgtccca 1212
QY 1081 gccagtgcggatatacaagcagatcagcttcagcttgcactctacacatcccttctcgtgtg 1140
Db 1213 gccagtgcggatatacaagcagatcagcttcagcttgcactctacacatcccttctcgtgtg 1272
QY 1141 gaggccaccccatggaccgctgtcctcctcctcgtgtgtggggggcgcacagagccgggca 1200
Db 1273 gaggccaccccatggaccgctgtcctcctcctcgtgtgtggggggcgcacagagccgggca 1332
QY 1201 gtttcctgtgtgagagagacatccaggggcagtgcaactcagtgaaagtggaatgc 1260
Db 1333 gtttcctgtgtgagagagacatccaggggcagtgcaactcagtgaaagtggaatgc 1392
QY 1261 atgtcacccctaaagatgccatcgcgcagccctgcacatttttgactgcccataatgg 1320
Db 1393 atgtcacccctaaagatgccatcgcgcagccctgcacatttttgactgcccataatgg 1452
QY 1321 ctggcacaggagtggtctcctgcagtgacatgtggccaggccctcagatcacogtgg 1380
Db 1453 ctggcacaggagtggtctcctgcagtgacatgtggccaggccctcagatcacogtgg 1512
QY 1381 gtcctctcagatccatcagatcaggaatgcacacagagcgctgtagcccaaaaacaaagccc 1440
Db 1513 gtcctctcagatccatcagatcaggaatgcacacagagcgctgtagcccaaaaacaaagccc 1572
QY 1441 cacataaagagggaatgcacatcgtaccctcctcctcgtataaacacaaagagaaattcca 1500
Db 1573 cacataaagagggaatgcacatcgtaccctcctcctcgtataaacacaaagagaaattcca 1632
QY 1501 gtcaggccaaagtgccatggttcaaaacagctcaagagctagaagaaggagctgtgtg 1560
Db 1633 gtcaggccaaagtgccatggttcaaaacagctcaagagctagaagaaggagctgtgtg 1692
QY 1561 tcaggagagccctcgtatccagagagcctggtgtgg 1597
Db 1693 tcaggagagccctcgtatccagagagcctggtgtgg 1729

RESULT 9
AAC78532
ID AAC78532 standard; cDNA; 1869 BP.
XX
AC AAC78532;
XX
DT 08-FEB-2001 (first entry)
XX
DE Human PR01071 (UNQ528) nucleotide sequence SEQ ID NO:300.
XX
KW Human; secreted protein; transmembrane protein; PRO; EST; cytostatic;
KW expressed sequence tag; detection; cancer; ss.
XX
OS Homo sapiens.
XX
PN WO200053756-A2.
XX
PD 14-SEP-2000.
XX
PF 18-FEB-2000; 2000WO-US04341.
XX
PR 08-MAR-1999; 99WO-US05028.
PR 12-MAR-1999; 99US-0123957.
PR 29-MAR-1999; 99US-0126773.
PR 21-APR-1999; 99US-0130232.
PR 28-APR-1999; 99US-0131445.
PR 14-MAY-1999; 99US-0134287.
PR 23-JUN-1999; 99US-0141037.
PR 26-JUL-1999; 99US-0145698.
PR 29-OCT-1999; 99US-0162506.

```

```

PR 30-NOV-1999; 99WO-US28313.
PR 02-DEC-1999; 99WO-US28551.
PR 02-DEC-1999; 99WO-US28551.
PR 16-DEC-1999; 99WO-US30095.
PR 30-DEC-1999; 99WO-US31243.
PR 30-DEC-1999; 99WO-US31274.
PR 05-JAN-2000; 2000WO-US00219.
PR 06-JAN-2000; 2000WO-US00277.
PR 06-JAN-2000; 2000WO-US00376.
XX
XX (GETH ) GENENTECH INC.
XX
PI Ashkenazi AJ, Baker KP, Botstein D, Desnoyers L, Eaton DL;
PI Ferrara N, Filvaroff E, Fong S, Gao W, Gerber H, Gerritsen ME;
PI Goddard A, Godowski PJ, Grimaldi CJ, Gurney AL, Hillan KJ;
PI Kljavin IJ, Kuo SS, Napier MA, Pan J, Paoni NF, Roy MA;
PI Shelton DL, Stewart TA, Tumas D, Williams PM, Wood WI;
XX
XX WPI; 2000-611443/58.
DR
DR P-PSDB; AAB44285.
XX
PT Novel PRO polypeptides and polynucleotides used in detection methods,
PT to target bioactive molecules to specific cells, and to modulate
PT cellular activities -
XX
XX Claim 2; Fig 119; 636pp; English.
XX
XX AAC78458 to AAC78599 represent polynucleotide and EST (expressed
XX sequence tag) sequences which encode secreted or transmembrane PRO
XX polypeptides. The PRO polynucleotides and polypeptides have cytostatic
XX activity. The polynucleotides and polypeptides can be used for detecting
XX the presence of PRO polypeptides in samples, for linking bioactive
XX molecules to cells and for modulating biological activities of cells,
XX using the polypeptides for specific targeting. The polypeptide targeting
XX can be used to kill the target cells, e.g. for the treatment of cancers.
XX The polypeptide pairs provide specific targeting of bioactive molecules
XX to cells. AAC78600 to AAC78987 represent PCR primers and probes used in
XX the isolation of the PRO polynucleotide sequences.
XX
XX Sequence 1869 BP; 490 A; 469 C; 480 G; 430 T; 0 other;

```

```

Query Match      29.7%; Score 1573; DB 21; Length 1869;
Best Local Similarity 99.1%; Pred. No. 0;
Matches 1582; Conservative 0; Mismatches 15; Indels 0; Gaps 0;

QY 1 atggaatgctgcgctcgggcaactcctcctgacacgtcctcttcttcttcttcttctcctc 60
Db 133 atggaatgctgcgctcgggcaactcctcctgacacgtcctcttcttcttcttcttctcctc 192

QY 61 ctgagttccagagcgcacgtccgagagagagagagagagagagagagagagagagagagagag 120
Db 193 ctgagttccagagcgcacgtccgagagagagagagagagagagagagagagagagagagagag 252

QY 121 ccattggagtgaatgctcagcacctcgggggtgggtgggtgggtgggtgggtgggtgggtgg 180
Db 253 ccattggagtgaatgctcagcacctcgggggtgggtgggtgggtgggtgggtgggtgggtgg 312

QY 181 ctgagcagcagagagctgtgaggaggaataatccgatacagaacacatgcagtaattgtgac 240
Db 313 ctgagcagcagagagctgtgaggaggaataatccgatacagaacacatgcagtaattgtgac 372

QY 241 tgcccaccagaagcaggtgatttcaggagcagcaatgctcagctcaataatgatgtcaag 300
Db 373 tgcccaccagaagcaggtgatttcaggagcagcaatgctcagctcaataatgatgtcaag 432

QY 301 caccatggccagttttatgaatgcttctcctgtgtcttaataaccctgcacacccatgttca 360
Db 433 caccatggccagttttatgaatgcttctcctgtgtcttaataaccctgcacacccatgttca 492

QY 361 ctcaagtgcacaagccaaagaaacacccctgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 420
Db 493 ctcaagtgcacaagccaaagaaacacccctgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 552

```

```
Qy 421 ggtacgctgtctatcacagaatcttttgatatgtgcatcagtggtttatgccaattgtt 480
Db 553 ggtacgctgtctatcacagaatcttttgatatgtgcatcagtggtttatgccaattgtt 612
Qy 481 ggtctgcatcacaccagctgggaagcacctgccaagaataaactgtggggtctgcaacgga 540
Db 613 ggcgcgataccacagctgggaagcacctgccaagaataaactgtggggtctgcaacgga 672
Qy 541 gatgggtccacctgccgctggtccgagggcaggtataataaccagctctccgcaacaaa 600
Db 673 gatgggtccacctgccgctggtccgagggcaggtataataaccagctctccgcaacaaa 732
Qy 601 tcgatatgatactggttgcaattccctatggaagtagacatatctgctgtcttaaaa 660
Db 733 tcgggatgatactggttgcaattccctatggaagtagacatatctgctgtcttaaaa 792
Qy 661 ggtctgatacttatatctggaaccacaaacccctccaggggactaaagggtgaaacagt 720
Db 793 ggtctgatacttatatctggaaccacaaacccctccaggggactaaagggtgaaacagt 852
Qy 721 ctcagctccacaggaacttctgtgacaattctagtggaacttcagaaatttcca 780
Db 853 ctcagctccacaggaacttctgtgacaattctagtggaacttcagaaatttcca 912
Qy 781 gacaaagagatactgagatggtggaaccactcacagcagatttcatgtcaagattcgt 840
Db 913 gacaaagagatactgagatggtggaaccactcacagcagatttcatgtcaagattcgt 972
Qy 841 aactcgggtccgctgacagtagacagtagacagtagacagtagacagtagacagtagac 900
Db 973 aactcgggtccgctgacagtagacagtagacagtagacagtagacagtagacagtagac 1032
Qy 901 tggagggagacagatttcttctgtctcagcaacctgtggaggaggttatcagctgcga 960
Db 1033 tggagggagacagatttcttctgtctcagcaacctgtggaggaggttatcagctgcga 1092
Qy 961 tcggctgagtgctacagatcgtgaggagcaaccgtgtggtgtgctgaccaatactgtca 1020
Db 1093 tcggctgagtgctacagatcgtgaggagcaaccgtgtggtgtgctgaccaatactgtca 1152
Qy 1021 taccagagaaacataaaaccccaacccagcttcagagtagtgaacttgatccttcca 1080
Db 1153 taccagagaaacataaaaccccaacccagcttcagagtagtgaacttgatccttcca 1212
Qy 1081 gccagtgacgatacaagcagatcagcttatgacctacacatcccttccctcgtgtg 1140
Db 1213 gccagtgacgatacaagcagatcagcttatgacctacacatcccttccctcgtgtg 1272
Qy 1141 gaggccaccccatgaccgctgtcctcctcgtgtggtgggggggcatccagagccgggca 1200
Db 1273 gaggccaccccatgaccgctgtcctcctcgtgtggtgggggggcatccagagccgggca 1332
Qy 1201 gtttctctgtgtagagagacatccaggggcatcacttcagtggaagagtgaatgc 1260
Db 1333 gtttctctgtgtagagagacatccaggggcatcacttcagtggaagagtgaatgc 1392
Qy 1261 atgtacacccctaaagtgcccatcgccagccctgcaacatttttgaactgcctaaatgg 1320
Db 1393 atgtacacccctaaagtgcccatcgccagccctgcaacatttttgaactgcctaaatgg 1452
Qy 1321 ctggcagagagtggtctcgtgacagtgacatgtggccagggcctcagataccgtgtg 1380
Db 1453 ctggcagagagtggtctcgtgacagtgacatgtggccagggcctcagataccgtgtg 1512
Qy 1381 gtctctgatacacaatcgaggaatgacacagggagcgtgtagcccaaaacaaagccc 1440
Db 1513 gtctctgatacacaatcgaggaatgacacagggagcgtgtagcccaaaacaaagccc 1572
Qy 1441 cacataaagaggaatgcatctgaccacatccctgctataaacccaaagagaacttcca 1500
Db 1573 cacataaagaggaatgcatctgaccacatccctgctataaacccaaagagaacttcca 1632
```

Qy 1501 gtcagaggccaagtgtccatgttcaaaacaaagctcaagagctagaagaaggagctctgtg 1560
Db 1633 gtcagaggccaagtgtccatgttcaaaacaaagctcaagagctagaagaaggagctctgtg 1692

Qy 1561 tcagaggagcctcttccatccagagagcctggtcgg 1597
Db 1693 tcagaggagcctctgaaagtgtaaaagcacagactg 1729

RESULT 10
AAS46142
ID AAS46142 standard; cDNA; 1869 BP.
XX
AC AAS46142;
DT 18-DEC-2001 (first entry)
XX
DE Human DNA encoding PRO polypeptide sequence #218.
XX
KW PRO polypeptide; mammal; tumour; cancer; human; cattle; horse; sheep; ss;
KW dog; cat; pig; goat; rabbit; tumour necrosis factor alpha; TNF-alpha;
KW blood; chondrocyte cell; cell proliferation; cell differentiation; colon;
KW adrenal; lung; breast; prostate; rectum; cervix; liver; genetic disorder;
KW PCR primer.
XX
OS Homo sapiens.
XX
PN WO200168848-A2.
XX
PD 20-SEP-2001.
XX
PF 28-FEB-2001; 2001WO-US06520.
XX
PR 01-MAR-2000; 2000WO-US05601.
PR 02-MAR-2000; 2000WO-US05841.
PR 03-MAR-2000; 2000US-187202P.
PR 06-MAR-2000; 2000US-186968P.
PR 14-MAR-2000; 2000US-189320P.
PR 14-MAR-2000; 2000US-189328P.
PR 15-MAR-2000; 2000WO-US06884.
PR 21-MAR-2000; 2000US-190828P.
PR 21-MAR-2000; 2000US-191007P.
PR 21-MAR-2000; 2000US-191048P.
PR 21-MAR-2000; 2000US-191314P.
PR 28-MAR-2000; 2000US-192655P.
PR 29-MAR-2000; 2000US-193032P.
PR 29-MAR-2000; 2000US-193053P.
PR 30-MAR-2000; 2000WO-US08439.
PR 04-APR-2000; 2000US-194449P.
PR 04-APR-2000; 2000US-194647P.
PR 11-APR-2000; 2000US-195975P.
PR 11-APR-2000; 2000US-196000P.
PR 11-APR-2000; 2000US-196187P.
PR 11-APR-2000; 2000US-196690P.
PR 11-APR-2000; 2000US-196820P.
PR 18-APR-2000; 2000US-198121P.
PR 18-APR-2000; 2000US-198585P.
PR 25-APR-2000; 2000US-199397P.
PR 25-APR-2000; 2000US-199550P.
PR 25-APR-2000; 2000US-199654P.
PR 03-MAY-2000; 2000US-201516P.
PR 17-MAY-2000; 2000WO-US13705.
PR 22-MAY-2000; 2000WO-US14042.
PR 30-MAY-2000; 2000WO-US14941.
PR 02-JUN-2000; 2000WO-US15264.
PR 05-JUN-2000; 2000US-209832P.
PR 28-JUL-2000; 2000WO-US20710.
PR 22-AUG-2000; 2000US-084488.
PR 22-AUG-2000; 2000WO-US23328.
PR 08-NOV-2000; 2000WO-US30952.
PR 01-DEC-2000; 2000WO-US32678.
PR 20-DEC-2000; 2000WO-US34956.
XX

(GETH) GENENTECH INC.

Baker KP, Chen J, Desnoyers L, Goddard A, Godowski PJ, Pan J, Smith V, Watanabe CK, Wood WI, Zhang Z, Gurney AL;

WPI; 2001-602746/68.

P-PSDB; AAU29241.

Novel nucleic acids encoding PRO polypeptides, used to diagnose the presence of tumours, such as prostate and breast tumours, in mammals and to screen for modulators of the compounds -

Claim 2; Fig 435; 774pp; English.

Sequences AAS45925-AAS46231 represent DNA molecules encoding and PCR primers for PRO polypeptides of the invention. The sequences of the primers can be used to detect the presence of a tumour in a mammal by comparing the level of expression of a PRO polypeptide in a test sample of cells from the animal and a control sample of normal cells, whereby a higher level of expression in the test sample indicates the presence of a tumour in the mammal. Mammals include dogs, cats, cattle, horses, sheep, pigs, goats and rabbits but are preferably human. The polypeptides can be used to stimulate tumour necrosis factor (TNF) alpha release from human blood, when contacted with it. A specific polypeptide can be used to stimulate the proliferation or differentiation of chondrocyte cells. The PRO proteins can be used to determine the presence of tumours and also susceptibility to tumour development, particularly adrenal, lung, colon, breast, prostate, rectal, cervical, or liver tumours, in mammalian subjects. The oligonucleotide probes specific for the PRO nucleic acids can be used for genetic analysis of individuals with genetic disorders.

Sequence 1869 BP; 490 A; 469 C; 480 G; 430 T; 0 other;

try Match	29.78;	Score 1573;	DB 22;	Length 1869;
Local Similarity	99.1%;	Pred. No. 0;		
Res 1582; Conservative	0;	Mismatches 15;	Indels 0;	Gaps 0;

1 atggaatgctgcgctgcgggcaactcctggcacactgctcctctttcttggtttctcgtc 60
|||
133 atggaatgctgcgctgcgggcaactcctggcacactgctcctctttcttggtttctcgtc 120

61 ctgagttccaggaccgcacgctccgaggagaccggacggcctatggatgctgggc 120
|||||
193 ctgagttccaggaccgcacgctccgaggagaccggacggcctatggatgctgggc 252

121 coatgagtgaaTgcTcacgcacctcggggtggggcctctactctctgagcgctgc 180
|||||
253 coatgagtgaaTgctcacgcacctcggggagggcctctactctctgagcgctgc 312

181 ctgagcagcaagctgtgaaggaagaataatccgatacagaaatcgagtaatgtggac 240
|||||
313 ctgagcagcaagctgtgtgaaggaagaataatccgatacagaaatcgagtaatgtggac 372

241 tgcccaccagaagcagggtgatttcgagctcagcaatgctcagctcataatgatgtcaag 300
|||||
373 tgcccaccagaagcagggtgatttcgagctcagcaatgctcagctcataatgatgtcaag 432

301 caccatggccagttttatgaatgggttcctggtgctaatgaacctgacaacccatgttca 360
|||||
433 caccatggccagttttatgaatgggttcctggtgctaatgaacctgacaacccatgttca 492

361 ctcaagtgccaaaggaacacccctggtgttgactagaccctaaaggtcttagat 420
|||||
493 ctcaagtgccaaaggaacacccctggtgttgactagaccctaaaggtcttagat 552

421 ggtacgcgttgctatcacagaatcttttgatatgtgcatacagtggttatgccaaattgtt 480
|||||
553 ggtacgcgttgctatcacagaatcttttgatatgtgcatacagtggttatgccaaattgtt 512

481 ggctgcgataccagctgggaagcacccgtcaaggaataactgtgggtctgcaacgga 540
|||||
513 ggctgcgataccagctgggaagcacccgtcaaggaataactgtgggtctgcaacgga 572

RESULT 11

AAF63447

ID AAF63447 standard; cDNA; 1803 BP.

XX AC

XX AC

XX AC

DT 14-MAY-2001 (first entry)

XX AC

DE Human ADAMTS-R1 cDNA sequence.

XX AC

ADAMTS-N; disintegrin; metalloprotease; thrombospondin type I motif;
 tumour cachexia; inflammation; dermatosparaxis; EDS-VIIC; angiogenesis;
 Ehlers-Danlos syndrome type VIIC; articular cartilage erosion; human;
 metastasis; embryogenesis; egg implantation; ADAMTS-R1; ss.

XX OS

XX Homo sapiens.

XX PN

XX WO200111074-A2.

XX PD

XX 15-FEB-2001.

XX PF

XX 03-AUG-2000; 2000WO-US21223.

XX PR

XX 06-AUG-1999; 99US-0369364.

XX XX

XX (CLEV-) CLEVELAND CLINIC FOUND.

XX PA

XX (APTE/) APTE S S.

XX PA

XX (HURS/) HURSKAINEN T L.

XX PA

XX (HIRO/) HIROHATA S.

XX XX

XX Apte SS, Hurskainen TL, Hirohata S;

XX PI

XX WPI; 2001-159978/16.

XX DR

XX P-PSDB; AAB72290.

XX PT

Murine and human 'A Disintegrin-like And Metalloprotease domain with
 Thrombospondin type I motifs' proteins and the nucleic acids encoding
 them, useful for treating e.g. tumours, inflammation and arthritis -

XX PS

XX Claim 15; Fig 11; 181pp; English.

XX CC

This invention relates to murine and human ADAMTS-N (A disintegrin-like
 and metalloprotease domain with thrombospondin type I motifs) proteins,
 designated ADAMTS-5, 6, 7, 8, 9, 10 and R1. Also included in the
 invention are cDNA sequences encoding the proteins, and antibodies
 specific for the proteins. The nucleic acid sequences and proteins may be
 used in the prevention, diagnosis and treatment of diseases associated
 with inappropriate ADAMTS-N expression. Disorders that may be treated
 using the nucleic acids, proteins and antibodies include, for example
 tumour cachexia, inflammation, dermatosparaxis in cattle or Ehlers-Danlos
 syndrome type VIIC (EDS-VIIC) in humans, erosion of articular cartilage
 in arthritic (both inflammatory and non-inflammatory) disease,
 angiogenesis, tumour growth and metastases, and they may also be used for
 controlling embryogenesis and implantation of fertilised eggs. The
 present sequence represents cDNA encoding human ADAMTS-R1.

XX XX

XX Sequence 1803 BP; 475 A; 460 C; 457 G; 411 T; 0 other;

Query Match

Best Local Similarity 29.7%; Score 1571.4; DB 22; Length 1803;

Matches 1581; Conservative 0; Mismatches 16; Indels 0; Gaps 0;

QY 1 atgaatgctgcctcgaggcaactcctggcacactgctctctcttctggtcttcgtc 60

|||||
 51 atgaatgctgcctcgaggcaactcctggcacactgctctctcttctggtcttcgtc 110

|||||
 61 ctgagttccagacgcacgcctccgagagacccggacggcctatggatgcctggggc 120

|||||
 111 ctgagttccagacgcacgcctccgagagacccggacggcctatggatgcctggggc 170

|||||
 121 ccatggagtgatgctacgcacctgcgggggtggggtcctctctctctgagggctgc 180

Db 171 ccatggagtgatgctacgcacctgcgggggtggggtccggtccactctctgagggctgc 230
 QY 181 ctgagcagcaagagctgtgaaggagaataatccgatacagacaacatgtaagtgtgac 240
 Db 231 ctgagcagcaagagctgtgaaggagaataatccgatacagacaacatgtaagtgtgac 290
 QY 241 tggccaccaggaagcaggtgatttcgagagctcagcaatgctcagctcaataatgtcaag 300
 Db 291 tggccaccaggaagcaggtgatttcgagagctcagcaatgctcagctcaataatgtcaag 350
 QY 301 caccatggccatttatgaatggcttcctgtcttaataagccctgacaacccatgtcca 360
 Db 351 caccatggccatttatgaatggcttcctgtcttaataagccctgacaacccatgtcca 410
 QY 361 ctcaagtgccaaagcacaacacccctggttggtaactagcaccctaaggtcttagat 420
 Db 411 ctcaagtgccaaagcacaacacccctggttggtaactagcaccctaaggtcttagat 470
 QY 421 ggtacgcgtgtgtatatacagaatcttggatattgtgcatcagtggttatgccaaattgt 480
 Db 471 ggtacgcgtgtgtatatacagaatcttggatattgtgcatcagtggttatgccaaattgt 530
 QY 481 ggtcgcgtatccacagctgggaagcacctcaaggaagataaactgtgggtctgcacacgga 540
 Db 531 ggtcgcgtatccacagctgggaagcacctcaaggaagataaactgtgggtctgcacacgga 590
 QY 541 gatgggtccacctgcgggtggtccgagggcagataataatcccagctctccgcaacaaa 600
 Db 591 gatgggtccacctgcgggtggtccgagggcagataataatcccagctctccgcaacaaa 650
 QY 601 tcggatgatactgtgttgcgaatccctatgggaagtagacatattgcctgtcttaaaa 660
 Db 651 tcggatgatactgtgttgcgaatccctatgggaagtagacatattgcctgtcttaaaa 710
 QY 661 ggtcctgatacttatctggaacacaaacccctcaggggactaaagtgaacacagt 720
 Db 711 ggtcctgatacttatctggaacacaaacccctcaggggactaaagtgaacacagt 770
 QY 721 ctacgtctccacaggaacttctctgttggaacaaattctagtgtggactccagaaattcca 780
 Db 771 ctacgtctccacaggaacttctctgttggaacaaattctagtgtggactccagaaattcca 830
 QY 781 gacaaagagatactgagaatggctggaccactcaacagatttcatgtccaagattcgt 840
 Db 831 gacaaagagatactgagaatggctggaccactcaacagatttcatgtccaagattcgt 890
 QY 841 aactcgggtccgcgtgacagtcacagtcacagttcttctatcaacccatcatccacga 900
 Db 891 aactcgggtccgcgtgacagtcacagtcacagttcttctatcaacccatcatccacga 950
 QY 901 tggaggagagcggatttcttctgtctcagcaacccctgtgagagaggttatcagctgaca 960
 Db 951 tggaggagagcggatttcttctgtctcagcaacccctgtgagagaggttatcagctgaca 1010
 QY 961 tcgctgagtgctacagatctgagagcaacccgtggtgtgctgaccaatactgtcaact 1020
 Db 1011 tcgctgagtgctacagatctgagagcaacccgtggtgtgctgaccaatactgtcaact 1070
 QY 1021 taccagagaacatcaaaccccaacccagcttcaggaggtgcaacttggtatcctgtcca 1080
 Db 1071 taccagagaacatcaaaccccaacccagcttcaggaggtgcaacttggtatcctgtcca 1130
 QY 1081 gccagtgacgatacaagcagatcatgcttatgactctaccatccctctcctcggtgg 1140
 Db 1131 gccagtgacgatacaagcagatcatgcttatgactctaccatccctctcctcggtgg 1190
 QY 1141 gaggccaccccatgaccgcgtgctcctcctcgtgtggggggggtatccagagcgggca 1200
 Db 1191 gaggccaccccatgaccgcgtgctcctcctcgtgtggggggggtatccagagcgggca 1250
 QY 1201 gttctcgtgtgtgagagagacatccaggggagatgcaacttcagtggaagagtggaatgc 1260

Db 1251 gttctctgtgtggaggaggacatccaggggcatgtcacttcagtgaagagtggaaatgc 1310
QY 1261 atgtacacccctaaagtgcoccatcgccagccctgcagacatttttgactgcccataatgg 1320
Db 1311 atgtacacccctaaagtgcoccatcgccagccctgcagacatttttgactgcccataatgg 1370
QY 1321 ctggcacagagtggtctcctcggtcacagtgacatgtggtgcccaggccctcagataccgtgtg 1380
Db 1371 ctggcacagagtggtctcctcggtcacagtgacatgtggtgcccaggccctcagataccgtgtg 1430
QY 1381 gtctctgcatcgaccatcgagggaatgcacacagagaggtgtgtagcccaaaaacaaagccc 1440
Db 1431 gtctctgcatcgaccatcgagggaatgcacacagagaggtgtgtagcccaaaaacaaagccc 1490
QY 1441 cacataaaagagaatgcacatcgtaaccactccctgtctatataaaacccaaagagaacttcca 1500
Db 1491 cacataaaagagaatgcacatcgtaaccactccctgtctatataaaacccaaagagaacttcca 1550
QY 1501 gtcgagggcgaagtgtccatgtgttcaacaagctcaagagctagaagagagctgtgtg 1560
Db 1551 gtcgagggcgaagtgtccatgtgttcaacaagctcaagagctagaagagagctgtgtg 1610
QY 1561 tcagagagagccctcgttcacatccagagggtggtcgg 1597
Db 1611 tcagagagagccctcgttaagtgttaaaagcacagactg 1647

RESULT 12

AAI66429

ID AAI66429 standard; cDNA; 1616 BP.

XX AAI66429;
XX ACXX 04-DEC-2001 (first entry)
DTXX Human thrombotic protein 46 coding sequence.
DEXX Human; thrombotic protein 46; cancer; HIV infection; gene therapy;
KW ss.XX Homo sapiens.
OSXX CN1300774-A.
PNXX 27-JUN-2001.
PDXX 22-DEC-1999; 99CN-0125685.
PFXX 22-DEC-1999; 99CN-0125685.
PRXX (BODE-) BODE GENE DEV CO LTD SHANGHAI.
PAXX Mao Y, Xie Y;
PIXX WPT; 2001-530462/59.
DRXX P-PSDB; AAG78679.
DRXX New human thrombotic protein 46 and polynucleotide encoding it, useful
PT for treating cancer and human immunodeficiency virus infection -
PTXX Claim 6; Page 24-25 (Disclosure); 32pp; Chinese.
PSXX The present invention provides the protein and coding sequences of human
CC thrombotic protein 46. The sequences can be used in the treatment of
CC cancer and HIV infection. The present sequence is the coding sequence of
CC the invention.XX Sequence 1616 BP; 386 A; 438 C; 426 G; 366 T; 0 other;
SQ

Query Match

Best Local Similarity 25.9%; Score 1369.2; DB 22; Length 1616;

Matches 1374; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

QY 638 gacatattcgcttctttaaaggctcctgacatctatatcttgaaacccaaacccctcc 697
Db 1 gacatattcgcttctttaaaggctcctgacatctatatcttgaaacccaaacccctcc 60
QY 698 agggactaaagggtgaaacagctctcagctccacaggaacttctctgtgacaaattcta 757
Db 61 agggactaaagggtgaaacagctcagctccacaggaacttctctgtgacaaattcta 120
QY 758 gtgtggacttcagaaatttcagacaaagagatctactgagaaatgctgagacactcacag 817
Db 121 gtgtggacttcagaaatttcagacaaagagatctactgagaaatgctgagacactcacag 180
QY 818 cagatttcattgtcaagattcgttaactcggctcgcgtgacagtagcagtcagttcatct 877
Db 181 cagatttcattgtcaagattcgttaactcggctcgcgtgacagtagcagtcagttcatct 240
QY 878 tctatcaaccatcatccaccgatggaggagacggattttcttctgtctcagaaacct 937
Db 241 tctatcaaccatcatccaccgatggaggagacggattttcttctgtctcagaaacct 300
QY 938 gtggaggaggttatcagctgacatcggctgagtgctacgatctgagagcaacogtgtg 997
Db 301 gtggaggaggttatcagctgacatcggctgagtgctacgatctgagagcaacogtgtg 360
QY 998 ttgctgaccaatctgtcactattaccagagaacatcaaaacccaaacccaaagcttcag 1057
Db 361 ttgctgaccaatctgtcactattaccagagaacatcaaaacccaaacccaaagcttcag 420
QY 1058 agtgcaacttggatccttctccagccagtgacggatacaagcagatcatccttatgacc 1117
Db 421 agtgcaacttggatccttctccagccagtgacggatacaagcagatcatccttatgacc 480
QY 1118 tctaccatcccttctcgtgtggaggagcccccagtgaccgctgtctctctcgtgtg 1177
Db 481 tctaccatcccttctcgtgtggaggagcccccagtgaccgctgtctctcgtgtg 540
QY 1178 gggggggcatccagagccggcaggttctctgtgtggaggagacatccaggggcatgca 1237
Db 541 gggggggcatccagagccggcaggttctctgtgtggaggagacatccaggggcatgca 600
QY 1238 cttcagtgaagagtggaaatgcattacacccctaaagtgcacccatgagtgccatcgccctgca 1297
Db 601 cttcagtgaagagtggaaatgcattacacccctaaagtgcacccatgagtgccatcgccctgca 660
QY 1298 acattttgactgcccataatggtggcagagtggtctcgtgcacagtgcacatgtg 1357
Db 661 acattttgactgcccataatggtggcagagtggtctcgtgcacagtgcacatgtg 720
QY 1358 gccagggctcagatcacctgtgtgtctctgcacagaccatgaggaatgcacacaggag 1417
Db 721 gccagggctcagatcacctgtgtgtctctgcacagaccatgaggaatgcacacaggag 780
QY 1418 gctgtagcccaaaaacaaagcccaataaaagaggaatgcacacccatccctgct 1477
Db 781 gctgtagcccaaaaacaaagcccaataaaagaggaatgcacacccatccctgct 840
QY 1478 ataaacccaaagagaacttccagtcgagggccaaagtgtgcatgggttcaaacagctcaag 1537
Db 841 ataaacccaaagagaacttccagtcgagggccaaagtgtgcatgggttcaaacagctcaag 900
QY 1538 agctagaagaagagctgtgtgtcagagagccctcgttcatccaggagcctgtgtg 1597
Db 901 agctagaagaagagctgtgtgtcagagagccctcgttcatccaggagcctgtgtg 960
QY 1598 cctgcacagtcacctgtgtgtggggaccagtgccgaatagtcaggtgcaggtgtcc 1657
Db 961 cctgcacagtcacctgtgtgtggggaccagtgccgaatagtcaggtgcaggtgtcc 1020
QY 1658 tgtctttctcagtcctgt 1717
Db 1021 tgtctttctcagtcctgt 1080

sequence is a DNA encoding human novel KIAA1233-like protein, NOV-2a. NOV-2a gene is located at chromosome 15.

SQ Sequence 7260 BP; 2022 A; 1593 C; 1811 G; 1834 T; 0 other;

Query Match 19.3%; Score 1023.4; DB 22; Length 7260;
Best Local Similarity 60.9%;
Pred. No. 2e-231;
Matches 1744; Conservative 0; Mismatches 1076; Indels 45; Gaps

QY	60	ctgagttccagaccgcacgctccgaggaggaccggagcgccctatgggatgctctgggg	119
Db	321	ccagacctcaagaacactggttcagatgaagacaagaatggcaactgggatgcttgggg	380
QY	120	cccatggagtgaaatgcttcacgcacctgcgggggtgggggctctactctctgagggcgtg	179
Db	381	cgactggagtgactgctcccgacctgtgggggaggagcatcatattctctcgaggatg	440
QY	180	cttgagcagcaagaagctgtgaaggagaataatccgatacagaaacatgcagtaattgga	239
Db	441	tttgactggaaggatgtgaaggcgaacattcggtaacaagacatcgagcaatcatga	500
QY	240	ctgcccaccagaagcaggtgatttccgagctcagcaatgctcagctcataatgatgtcaa	299
Db	501	ctgcccctccagatgcagaagatttcagagccgcagtgctcagcctcaaatgatgtcaa	560
QY	300	gcaccatggccagtttatgaatggcttctgtctaatgaacctcgacaacccatgttcc	359
Db	561	gtatcaagggttactatgaatggcttccacgataataatgcttgcgtgcccggtgac	620
QY	360	actcaagtgcacaagcccaaaagcaacccctggttgttgtaactgacacctaaggctctaga	419
Db	621	actcaagtgtcatgcacaaggcaaaacttgggtggagctggcgaactaaagttactgga	680
QY	420	tggtaacggtgtgtatcacagaatttttgatatgtgcatacagtggttttatgccaaattgt	479
Db	681	tggaaactcgtgtgaacacggactccttggacatgtgtatcagtggcactctcagggcagt	740
QY	480	tggctgcgatccacagctgggaagcaccgtcaaggaagataaacttgggggtctgcaacgg	539
Db	741	gggctgcgactcggcaactgggaagcaatgcccaaggaggacaactggaggtctgtgcgg	800
QY	540	agatgggtccaactgcggctggctgcgaggcagtgataaataccagctctccgcgaaccaa	599
Db	801	cgatggctccaactgcgaggttgtacggggacaatacaaatcacagtttctcttgaaaa	860
QY	600	atcggatgtacttgggttgcaattccctatggaagtagacatatcgcccttgttctaaa	659
Db	861	aagagaagaaaaatgtaatgtcttcccttgggaagtcgaagtgagaaattcacagtga	920
QY	660	aggctcgtgacattatcttgaaaccaaaccctccaggggactaaagttggaagaaacag	719
Db	921	aggacctgccacctttatgaaataaaaaacacttcaaggaaacgaaggagaaacacag	980
QY	720	tctcagctccaagaaacttccctgtggacaattctagtgtggacttccagaaatttccc	779
Db	981	ctttaacagccccggcgtttgtcgtagaacaacacacagtggaatttcagaggggctc	1040
QY	780	agacaagaagatactgagaatggctggaccactcaacagacagatttcatgtcaagattccg	839
Db	1041	cgagagggcaaaactttaagattccaggacctctgatttcttcatcttccacagaccag	1100
QY	840	taactcgggctccgctgacagtcacagtcacgttccatctctatcaaccatcattccaccg	899
Db	1101	gtacactgcagccaaagacagcgtggttcagttcttcttcttaccagcccatcagtcata	1160
QY	900	atggaagggagacggtatttcttctgtctcagcaacctgtggaggaggttatcagctgac	959
Db	1161	gtggagacaaactgacttcttccctgcacigtgcgtgtgcgtgtggaggaggttatcagctcaa	1220
QY	960	atcggcttgagtgctacgactctgaggaagcaacccgtgtggttctgacccaatactgtcacta	1019
Db	1221	ttctcgtcgaatgtggataccgcttgaaagagggtagttcctgaccattattctgacta	1280

Wed Jul 24 11:35:34 2002

```

QY 2100 gatgaatgaacagctcatcctggctgatgagctgtgtgcccagcccaagcccagcacggt 2159
Db 2333 -----cagggagaccctgcccctcctgagggagtgccgagatgaaagcccacatgcttt 2387
QY 2160 gcaagcttgaaccgttttaattgccccccagcctggtaacctgcacagtgccagccgtg 2219
Db 2388 acaagcatgcaatcagtttgaactgcccctcctgctggcgcacattgaagaatggcagcgtg 2447
QY 2220 ttccagaacgtgtgcccgggggtgttcagaacagtgaggttcttggcaagcagcgcgtggc 2279
Db 2448 ttccaggactgtggcgggggaaactcagacaagagaagagtcacactgtcggcagctgctaac 2507
QY 2280 tgatggcagcttccctggagcttctcagagactccttctgttcagcttcaaaaacctgctgcca 2339
Db 2508 ggatggcagcttttgaatctcagatgaattgigccaaaggaccacaaggcatcgtctca 2567
QY 2340 gcaagcatgcaagaagatgactgtcccagcaggtggttctctcagactggacagagtg 2399
Db 2568 caagtctgtgccaggacagactgtcctccacatttagctgtgggagactgggtcgaagtg 2627
QY 2400 ttccacaagctgcggggaagcaccagactcgaagcgccatttgcgcaaaagatgctgaa 2459
Db 2628 ttctgcagttgtgtgtgttggaatccagagaagaagcaggtgtgtcaaggctggcagc 2687
QY 2460 aacccgctctcaacggttgcgaattccacccctgtgcccgccttgccttctctctc 2519
Db 2688 caaagtcggcgcatcccctcagtgagatgagtgagggatctaccaggggttccctct 2747
QY 2520 catcaggccctgtatctggc-----aacctgtgcaaggcccgggcgccatc 2567
Db 2748 tgtaagatcttgcagatgctgagtgagtgcaagtaaaatcaaatcagagatgaagacaaaact 2807
QY 2568 caggaagcacagcccacatcgcggccgaggaaggtctacatccagactgcagggca 2627
Db 2808 tggtagcaggggtcccgagatcctcagttccagagaggtctacattcagacaagggaaga 2867
QY 2628 gaggaagctgcaactctgtgtgggggcttgcctcactctgcctccccaagacggcggtggt 2687
Db 2868 gaagcgtattaaacctgacctggttagcagagcctatttgcctcccaacacatccgtgat 2927
QY 2688 gctgcgctgcccgcgcgcaggttccgcaagccctcactcactgggagagaagacggcca 2747
Db 2928 tattaagtgcctcgtgcgagattccagaaatctctgatccagtgaggagagatggccg 2987
QY 2748 gcacctcatcagctcgacgcacgtlccaggtggccctctcggtctatctcaagatcccaccg 2807
Db 2988 ttgcctgcagaactccaaacggcttggcatcccaagtcaggctcactaaaaatccacgg 3047
QY 2808 cctcaagccctcgatgcagggctctacacctgtcagcggccggccggggagacatt 2867
Db 3048 tcttgcctcccgacatcggtgtaccggtgcatgtcaggtctgtcacagggaacagt 3107
QY 2868 tgtgattaaagctcatcgaggcaaccgcaagctcgtgcccggcc 2912
Db 3108 tgtgtcaagctcattgtgactgacaaccggctcactgcagcgccc 3152

```

Search completed: July 24, 2002, 02:24:47
Job time: 11016 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 24, 2002, 04:22:27 ; Search time 25.27 Seconds
(without alignments)
2699.796 Million cell updates/sec

Title: US-10-044-807-2

Perfect score: 9588

Sequence: 1 MECCRRATPGTLLFLALL.....LKLQLSQPKSCCGTCGKA 1762

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 105224 seqs, 38719550 residues

Total number of hits satisfying chosen parameters: 105224

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_40.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	972	10.1	1629	1	Q9p2n4 homo sapien
2	894	9.3	1593	1	P58397 homo sapien
3	807	8.4	1077	1	Q9h324 homo sapien
4	651	6.8	1201	1	O15072 homo sapien
5	639.5	6.7	1205	1	P79331 b adamts-2
6	613	6.4	967	1	Q9wug1 rattus norv
7	608.5	6.3	1211	1	O95450 h adamts-2
8	597	6.2	968	1	P97857 mus musculu
9	584.5	6.1	967	1	O9uh18 homo sapien
10	564.5	5.9	450	1	P58459 mus musculu
11	506	5.3	860	1	Q9uhp5 homo sapien
12	495.5	5.2	890	1	Q9up79 homo sapien
13	474.5	4.9	905	1	P57110 mus musculu
14	469.5	4.9	930	1	Q9r001 mus musculu
15	462.5	4.8	930	1	O9una0 homo sapien
16	422	4.4	837	1	O75173 homo sapien
17	407.5	4.3	997	1	Q9ukp4 homo sapien
18	401	4.2	630	1	Q9esp7 rattus norv
19	376	3.9	4393	1	P98160 homo sapien
20	332.5	3.5	1906	1	O15746 homo sapien
21	328	3.4	1914	1	Q05695 rattus norv
22	326	3.4	3707	1	O14514 homo sapien
23	320.5	3.3	1259	1	O60241 homo sapien
24	317	3.3	1584	1	P43146 homo sapien
25	312.5	3.3	1572	1	P11627 mus musculu
26	311.5	3.2	469	1	P32004 homo sapien
27	311.5	3.2	1447	1	P35331 gallus gall
28	307.5	3.2	1260	1	P70211 mus musculu
29	299	3.1	1257	1	P11680 mus musculu
30	299	3.1	1284	1	P20241 drosophila
31	297.5	3.1	1447	1	
32	290.5	3.0	437	1	
33	289.5	3.0	1302	1	

ALIGNMENTS

RESULT 1
 ID AT9_HUMAN STANDARD; PRT; 1629 AA.
 AC Q9P2N4; Q9NR29;
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DE 01-MAR-2002 (Rel. 41, Last annotation update)
 DE ADAMTS-9 precursor (EC 3.4.24.-) (A disintegrin and metalloproteinase with thrombospondin motifs 9) (ADAM-TS 9) (ADAM-TS9).
 GN ADAMTS9 OR KIAA1312.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A. (SHORT ISOFORM).
 RC TISSUE=Fetal;
 RX MEDLINE=203961138; PubMed=10936055;
 RA Clark M.E., Kellner G.S., Turbeville L.A., Boyer A., Arden K.A., Maki R.A.;
 RT "ADAMTS 9, a novel member of the ADAM-TS/Metallospondin gene family";
 RL Genomics 67:343-350(2000).
 RN [2]
 SEQUENCE OF 159-1629 FROM N.A. (LONG ISOFORM).
 RC TISSUE=Brain;
 RX MEDLINE=20181126; PubMed=10718198;
 RA Nagase T., Kikuno R., Ishikawa K.-I., Hirose M., Ohara O.;
 RT "Prediction of the coding sequences of unidentified human genes. XVI. The complete sequences of 150 new cDNA clones from brain which code for large proteins in vitro.";
 RL DNA Res. 7:65-73(2000).
 CC -! COFACTOR: BINDS 1 ZINC ION (BY SIMILARITY).
 CC -! SUBCELLULAR LOCATION: SECRETED. ASSOCIATED WITH THE EXTRACELLULAR MATRIX (BY SIMILARITY).
 CC -! ALTERNATIVE PRODUCTS: 2 ISOFORMS; A LONG FORM (SHOWN HERE) AND A SHORT FORM; MAY BE PRODUCED BY ALTERNATIVE SPLICING.
 CC -! TISSUE SPECIFICITY: HIGHLY EXPRESSED IN ALL FETAL TISSUES. EXPRESSED SLIGHTLY IN ADULT OVARY, PANCREAS, HEART, KIDNEY, LUNG, PLACENTA. ALSO DETECTED IN SPINAL CORD AND BRAIN. NOT DETECTED IN COLON, SMALL INTESTINE, TESTIS, LIVER, SKELETAL MUSCLE, SPLEEN OR THYMUS.
 CC -! DOMAIN: THE SPACER DOMAIN AND THE TSP TYPE-1 DOMAINS ARE IMPORTANT FOR A TIGHT INTERACTION WITH THE EXTRACELLULAR MATRIX (BY SIMILARITY).
 CC -! PTM: THE PRECURSOR IS CLEAVED BY A URIN ENDOPEPTIDASE (BY SIMILARITY).
 CC -! SIMILARITY: BELONGS TO PEPTIDASE FAMILY M12B.
 CC -! SIMILARITY: CONTAINS 1 DISINTEGRIN-LIKE DOMAIN.
 CC -! SIMILARITY: CONTAINS 11 TSP TYPE-1 DOMAINS.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial

34 282 2.9 1077 1 SM5A_MOUSE Q62217 mus musculu
 35 279 2.9 1074 1 SM5A_HUMAN Q13591 homo sapien
 36 278.5 2.9 5376 1 ZAN_MOUSE O86799 mus musculu
 37 272.5 2.8 1040 1 AXOL_HUMAN Q02246 homo sapien
 38 267 2.8 2012 1 DSCA_HUMAN O60469 homo sapien
 39 265 2.8 1051 1 PTK7_CHICK Q91048 gallus gall
 40 264 2.8 807 1 FSPQ_RAT P35446 rattus norv
 41 258.5 2.7 1040 1 AXOL_RAT P22063 rattus norv
 42 257 2.7 470 1 PROP_CAVPO Q64181 cavia porce
 43 253.5 2.6 1093 1 SM5B_MOUSE O60519 mus musculu
 44 252.5 2.6 1170 1 TSP1_BOVIN Q28178 bos taurus
 45 250.5 2.6 1897 1 PTPF_HUMAN P10586 homo sapien

entities requires a license agreement (See <http://www.isb-sib.ch/announcement/> or send an email to license@isb-sib.ch).

CC	EMBL; AF261918; AAF89106.1; -	
DR	EMBL; AB037733; BAA92550.1; -	
DR	MIM; 605421; -	
DR	InterPro; IPR001762; Disintegrin.	
DR	InterPro; IPR02870; IPR_M12B_propcp.	
DR	InterPro; IPR001590; Reprolysin.	
DR	InterPro; IPR000884; TSP1.	
DR	InterPro; IPR000130; Zn_MTpeptdse.	
DR	pfam; PF01562; Pep_M12B_propcp; 1.	
DR	pfam; PF01421; Reprolysin; 1.	
DR	pfam; PF00090; tsp-1; 11.	
DR	SMART; SM00209; TSP1; 12.	
DR	PROSITE; PS0215; ADAM_MEPRO; 1.	
DR	PROSITE; PS00427; DISINTEGRIN_1; FALSE_NEG.	
DR	PROSITE; PS00092; TSP1; 9.	
DR	PROSITE; PS00142; ZINC_PROTEASE; 1.	
DR	Hydrolase; Metalloprotease; Zinc; Signal; Glycoprotein; Zymogen;	
KW	Repeat; Extracellular matrix; Alternative splicing.	
KW	POTENTIAL.	
FT	SIGNAL	1 18
FT	PROPEP	19 287
FT	CHAIN	288 1629
FT	DOMAIN	509 587
FT	DOMAIN	589 642
FT	DOMAIN	645 752
FT	DOMAIN	753 880
FT	DOMAIN	999 1053
FT	DOMAIN	1056 1108
FT	DOMAIN	1111 1156
FT	DOMAIN	1184 1239
FT	DOMAIN	1240 1295
FT	DOMAIN	1320 1383
FT	DOMAIN	1386 1439
FT	DOMAIN	1445 1498
FT	DOMAIN	1501 1554
FT	DOMAIN	1562 1612
FT	DOMAIN	88 96
FT	SITE	223 223
FT	METAL	434 434
FT	ACT_SITE	435 435
FT	METAL	438 438
FT	METAL	444 444
FT	CARBOHYD	112 112
FT	CARBOHYD	135 135
FT	CARBOHYD	171 171
FT	CARBOHYD	271 271
FT	CARBOHYD	749 749
FT	CARBOHYD	840 840
FT	CARBOHYD	1213 1213
FT	CARBOHYD	1267 1267
FT	VARSPLIC	1064 1072
FT	VARSPLIC	1073 1629
FT	CONFLICT	367 367
FT	SEQUENCE	1629 AA; 182649 MW; C1C4CEFF58B941F CRC64;
SO		F -> L (IN REF. 1).

Query Match 10.1%; Score 972; DB 1; Length 1629;
Best Local Similarity 25.4%; Pred. No. 4.4e-52;
Matches 306; Conservative 152; Mismatches 391; Indels 354; Gaps 55;

[illegible]

FT	DOMAIN	1426	1471	TSP TYPE-1 7.
FT	DOMAIN	1426	1471	TSP TYPE-1 8.
FT	DOMAIN	302	305	POLY-GLU.
FT	SITE	208	208	CYSTEINE SWITCH (POTENTIAL).
FT	METAL	392	392	ZINC (CATALYTIC) (BY SIMILARITY).
FT	ACT SITE	393	393	BY SIMILARITY.
FT	METAL	396	396	ZINC (CATALYTIC) (BY SIMILARITY).
FT	METAL	402	402	ZINC (CATALYTIC) (BY SIMILARITY).
FT	CARBOHYD	105	105	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	125	125	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	215	215	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	485	485	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	685	685	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	790	790	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	951	951	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	1104	1104	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	1275	1275	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	1300	1300	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	1320	1320	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	1371	1371	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	1378	1378	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	1503	1503	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	SEQUENCE	1593 AA;	177545 MW;	07F9F48E63B083A3 CRC64;
Query Match				
Best Local Similarity 9.3%; Score 894; DB 1; Length 1593;				
Matches 358; Conservative 169; Mismatches 451; Indels 812; Gaps				
QY	27	RSEEDRDLWDAMPWSECSTCTGGGASYSLRRC-----LSSKSCBGRNIRVYTCSNVD	80	
DB	536	KKPESTPGGWRGSPWHSRCSCTGAGVQSAERLCNNPEPKFGKYCTGEKRYKLCNVHP	595	
QY	81	CPEAGDFRAQOCSAHDVXKHQGYFELPVSNDDNPCLSKCOAKGTTLLVWLAPKVL	140	
DB	596	CRSEAPTFRMOCSEFDTPVYKNELYHWFPIFN-PAHPCELYCRPIDGQFSKMLDAVID	654	
QY	141	GTRCYT--ESLDMCISGLCOIVGCDHOLGSTVEDNGVCGNGDGSCTRLVRGYKSKLSA	198	
DB	655	GTPCFEGGSRNVCINGICAKWGCDVEIDSNATEDRCGVCGLDGSSCOTVRKMFKQEGS	714	
QY	199	TKSDDTVVAIPYGRHRLVLYKGPDLHLYLETKTLOGTKGENSLSTGTFLVDNSSVDFOK	258	
DB	715	GYVD--IGLIPKRGARDIRV-----EIEGAGNFLAIRS-----E 746		
QY	259	FPDKETLRWAGPLTADFIVKIRNSGADSTVQIFQPIIHRWRETDFFPCSTCGGGYQ	318	
DB	747	DPKYYL-----NGG-----FTIOW-----NGNYK 766		
QY	319	LTSAECYDLRSNRVADQYCHYYPENIKPKPKLQECNLDPASPADGKYQIMPYDLYHPLP	378	
DB	767	LACT-----VFQYDRKGDL 781		
QY	379	RWEATPTWACSSCGGGIOSRAVSCVEEDIQGHVTSVEEWKCMYTPKMPIAQCNTFDPC	438	
DB	782	KLMAT-----GPTNESV----- 793		
QY	439	KWLAEQSPCTVTCGGGLRVVVLCDHRGMHTGCGSPKTKPHIKECIVPTPYKPKK 498		
DB	794	WIQ-----LRFQV-----TNFGIKYEYTI----- 812		
QY	499	LPVEAKLPWFOAKELEGAAVSEEPSTPEAWSACTVTTCGVGTQVIRVRCVLLSFSQS	558	
DB	813	----- 812		
QY	559	VADLPIDCEGPKPASORACYAGPCSGEIPENPDETGLFGGLQDFDLYDMWYEGFTK	618	
DB	813	-----QKGLD---NDVEQMYFQWYGHWTE 834		
QY	619	CSCEGGGVQOEA VVSLNKTOTREPAENLCVTSRRPPQLLKSNLDDPCPARWEIKWSPC	678	
DB	835	CSVTGCTGIRROTANCI-KKGRGMVKATFCDPETQPNRGQKKCHEKACPPRWAGEWAC	893	

```
QY 679 SUTCGVGLQTRDFVCSHLLSREMNFTVILADELCROPKESTVOACNRFNCPNPPAWYPAQWQ 738
Db 894 SATCG-----PH-----900
QY 739 PCSRTCGGVQKREVLCQKQRMADGSLPELTFCS-ASKPACQACKKDD-CPSEWLLSD 796
Db 901 -----GKRRVLCIOTWVSD-QUALPPDQCHLLPKTKLLSCNRDILCPDSMTWGN 951
QY 797 WTECSTSCGEGTQTRSAICRKMKTGLSTVWVNSTLCPPLPFSSSIRPCMILATCARPGRS 856
Db 952 WSECVSVCGGVIRSVTCAKNUDE-----PCDVTKPNRSALCG-----991
QY 857 THSPHIAARKVYIOTRRQRKLIHFVGGFAYLLPNTAVVLRCP-ARRVRKPKLITWEKDG 915
Db 992 -----LQCFSSRRVLKP-----1004
QY 916 QHLISSTHVAPFGYKLIHRLKPSDAGVYTCAGPAREHFVILKIGGNKLVARPLSPR 975
Db 1005 -----1004
QY 976 SEEVLAGRKGPKALQTHKHQNGIFSGSKAEKRGLAANPGSRYDDLVSRLLQGGWP 1035
Db 1005 -----NKGTSNGKNPPTLKPVPPTSR-----PRMLTTPTGP 1037
QY 1036 GELLASWEAODSAERNWTSEE-DPGAEOVLLHLPFTVTEQRLLDILGNLSQOPELRD 1094
Db 1038 ESMSTPTAISPTASPTASKEGDLGGK-----WQDSSTQPE-----1074
QY 1095 LYSKHLVA--QLAEIFRSHLEHODVLLKPSERRTSPVTLSPHKHVSFGFSSLSRTSTGD 1152
Db 1075 LSSKYLSTGSTSOPILTS-----QSLSIQPSEENVSSDTGP-----TSEGLVATTISGS 1126
QY 1153 AGGSRPRPKP-----TILRKISAAQLSASEVVTGLGTVALASGTVLLHCEAIGH 1207
Db 1127 GLSSRNPIWVTPPTTLTK-----GPEMEHSG-----1157
QY 1208 PRPTISWARGEEVQSRDRIILQDDSLQILAPVEADVGYTCNATNALGYDSVSIATVL 1267
Db 1158 -----SGEERE-----QPEDK-----DESNPVTWTKIRVP-----1182
QY 1268 AGPLVKTSRMTVINTKPAVTVDIGSTIKTVQGVNVTINQVAGVPEAEVTFNRKSKL 1327
Db 1183 GNDAPVESTEMPL-----APLTPDL-----SRESWPPPTST- 1214
QY 1328 GSPHLLHEGSLLLTVNSSDDQGLYSCRAANLHGLTESTQLLLDPPQVPTQLEDIRALL 1387
Db 1215 -----VMEGLL-----PSORPTTSE-----1229
QY 1388 AATG-PNLPSVLTPIGTQIVLDGNSALLGCCPIKG--HPVP-----NITW 1430
Db 1230 --TGTPRVGNVTE-----KPANTLL--PLGGDHQPEPSGKTANRNLKLPNNNQ 1276
QY 1431 FHGQPTVATGLTHHILAAGQILQVANLSGSGSEFSCLAQNEAGVLMOKASLVQIDYW 1490
Db 1277 TKSEPVLTEDAT-SLITTEGFLNASNKYOLTNHGSA-----H 1315
QY 1491 WSDVRLATCASGNRGVQVQPRLLKCLLNSTEVNPAHCAKGVKRPVAVQPI-ACNRRDCPSRW 1549
Db 1316 WIVGNWSECSTTCG-LGAYWKRVEE--TTQMD-SDCAAIIQRP--DPAKRCHLRPC-AGW 1367
QY 1550 MVTWSACTSCGGGVOTRRVTC-----OKLK-----ASGISTPVSNDCQTVAKRPV 1597
Db 1368 KVGWNSKSRNCSGGFKIREIQCVDSRDHNLRFHFHCOFLAGIPPLUS-----1415
QY 1598 DTQACNOLCVWEAFSSWGQNGPCIGPHLAVQHRQVFCQTRDGTILFSEQCSALPRVPS 1657
Db 1416 --MSCNPEPCAWQVPEWQSCRSRSCGG--GVQERGVEE-----PGGLCDWTKRETS 1462
QY 1658 TQNCWSEACSVHWRVSVLWLTCTATCGNFGQSRVVECVHARTKANVPEH--LCSWGPRPA 1715
Db 1463 TMSCNELHC-CHWATGNWDLGSTSCGG-GFOKRIVQCVPSSEGNKTEODQCLCHKPRPP 1520
QY 1716 NWQRNITPCE--NMEC---RDITRYCEKVKQLKLCQLSQFKSRCCTGC 1759
```

```
Db 1521 EFKKCQQACKKSADLLCTKDKLSASFQCLTKAMKKCSVFTVRAECFCSC 1570
RESULT 3
AT10_HUMAN STANDARD; PRT; 1077 AA.
ID AT10_HUMAN Q9H324;
AC Q9H324;
DT 01-MAR-2002 (Rel. 41, Created)
DT 01-MAR-2002 (Rel. 41, Last sequence update)
DT 01-MAR-2002 (Rel. 41, Last annotation update)
DE ADAMTS-10 precursor (EC 3.4.24.-) (A disintegrin and metalloproteinase
DE with thrombospondin motifs 10) (ADAM-TS 10) (ADAM-TS10) (Fragment).
GN ADAMTS10.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Apte S.S.;
RT "ADAM-TS10: a novel member of the ADAM-TS family containing multiple
RT thrombospondin type I repeats.";
RL Submitted (JUN-1999) to the EMBL/GenBank/DBJ databases.
CC -|- COPACTOR: BINDS 1 ZINC ION (BY SIMILARITY).
CC -|- SUBCELLULAR LOCATION: SECRETED, ASSOCIATED WITH THE EXTRACELLULAR
CC MATRIX (BY SIMILARITY).
CC -|- DOMAIN: THE SPACER DOMAIN AND THE TSP TYPE 1 DOMAINS ARE IMPORTANT
CC FOR A TIGHT INTERACTION WITH THE EXTRACELLULAR MATRIX (By
CC similarity).
CC -|- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M12B.
CC -|- SIMILARITY: CONTAINS 1 DISINTEGRIN-LIKE DOMAIN.
CC -|- SIMILARITY: CONTAINS 5 TSP TYPE-1 DOMAINS.
CC -----
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement. Usage by and for commercial
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: AF163762; AAC35563.1;
DR InterPro: IPR002870; Pep_M12B_propep.
DR InterPro: IPR001590; Repolysin.
DR InterPro: IPR000884; TSP1.
DR InterPro: IPR000130; Zn_MTpeptdse.
DR Pfam; PF01562; Pep_M12B_propep; 1.
DR Pfam; PF01421; Repolysin; 1.
DR Pfam; PF00090; tsp_1; 5.
DR SMART; SM00209; TSP1; 5.
DR PROSITE; PS50215; ADAM_MEPRO; 1.
DR PROSITE; PS50092; TSP1; 2.
DR PROSITE; PS00142; ZINC_PROTEASE; 1.
DR PROSITE; PS00427; DISINTEGRIN_1; FALSE_NEG.
KW Hydrolase; Metalloprotease; zinc; Glycoprotein; zymogen;
KW Repeat; Extracellular matrix.
FT CHAIN 1 207 BY SIMILARITY.
FT PROPEP <1 207 ADAMTS-10.
FT METAL 208 1077 ZINC (CATALYTIC) (BY SIMILARITY).
FT ACT_SITE 366 366 ZINC (CATALYTIC).
FT METAL 367 367 ZINC (CATALYTIC) (BY SIMILARITY).
FT METAL 370 370 ZINC (CATALYTIC) (BY SIMILARITY).
FT METAL 376 376 ZINC (CATALYTIC) (BY SIMILARITY).
FT DOMAIN 434 520 DISINTEGRIN-LIKE.
FT DOMAIN 578 679 SPACER.
FT DOMAIN 680 802 TSP TYPE-1 1.
FT DOMAIN 802 802 TSP TYPE-1 2.
FT DOMAIN 860 918 TSP TYPE-1 3.
FT DOMAIN 922 976 TSP TYPE-1 4.
FT DOMAIN 981 1031 TSP TYPE-1 5.
FT CARBOHYD 64 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT 64
```

FT	CARBOHYD	196	196	N-LINKED (GLCNAC. . .)	(POTENTIAL).
FT	CARBOHYD	297	297	N-LINKED (GLCNAC. . .)	(POTENTIAL).
FT	CARBOHYD	714	714	N-LINKED (GLCNAC. . .)	(POTENTIAL).
FT	CARBOHYD	769	769	N-LINKED (GLCNAC. . .)	(POTENTIAL).
FT	CARBOHYD	866	866	N-LINKED (GLCNAC. . .)	(POTENTIAL).
SQ	SEQUENCE	1077	AA; 118072 MW; 3914DE18DCBFF587	CRC64;	
Query Match					
Best Local Similarity 32.9%; Score 807; DB 1; Length 1077;					
Matches 197; Conservative 71; Mismatches 224; Indels 106; Gaps 21;					
QY	30	EDRDGLDAMGPWMSRCSRTCGGASYSURRC-----LSSRSCGRNRYRTCSNVDCPP	83		
DB	518	EGVDGAMPWTWGDCTCTCGGVSSSRHCDSPRPTGGKYICLGERRRHRSNTDDC	577		
QY	84	EAGFRAQCCSAHNDKHHGFYEWLPVSNPDNPNCSLKQAKGTTLVVVLAPKVLDTGR	143		
DB	578	GSQDFREVQCSEFDSIPRFGFKYKTVRGVGVKACSLTSLAEGNFYTERAAVVDGIP	637		
QY	144	CYTESLDMCISGLQIVCGDHLQSTYKEDNGVCGVNGDSTCRIVRGYKKSOLSAKSD	203		
DB	638	CRPTVDICVSGECKHVGCDRLVGSDDREDKRCVGGDGSACETIEGVF-SPASPGAGYE	696		
QY	204	TVVAIPYXGSRHI-----RLVLKGPDLHLYETKTLOGTKFNSLSSTCTFLVDNSS	253		
DB	697	DVWIPKGSVHIFTDLNLSLSHLAKG-DOESLLEGLPGTPQPHRLPLAGT-----T	749		
QY	254	VDFQKFPDK-EILRMAGPLTADFIKIRNGSGADSTVQFIYQPII-----HRWRETF	306		
DB	750	FQLRQGPDOVQSLPAGLFINASLLVMVL-ARTELPALRYRNAPIALRSLPPYSHWYAPW	808		
QY	307	FPCSATCGGVYLTSAECYDLRSNRVADQYCHYPENIKPKLQE-CNLDPCPASDGY	365		
DB	809	TKSAQCAGGSQVQAVECRNOLDSSAVAPHYCSAHSK-----LPRQRACNTEPCP	859		
QY	366	KQIMPYDLYHPLPWEATPWTACSSCGGIGQSRVAVSCVEEDIQHVTISVEW-----KCM	421		
DB	860	-----PDWVGNLSLSCSDAGVRSRVVC-----ORRYSAAEKALDSDAC	902		
QY	422	YTPKMPATQCNITDC-PKWLAQEWSPCTVTCGGGLRYRVVLC---IDHRG-MHTGGCSPK	477		
DB	903	POPRPPVLEACHGTPCPPEAAALDWSBCTPSCGPLRHRVVLCSDHRATLPPAHGSPA	962		
QY	478	TKPHKEECIVTPCYKPEKLPVEAKLPWFQKQAELEGAANVEEPSFIPEAWSACTVT	537		
DB	963	AKPPATMRCNL-----RRCP-----PARWVAGEWGECSAQ	992		
QY	538	CGVGTQVRIYRCQVLLSFSQSVADLPIDEC-EGPKPASORACYAGPCSGEIPFENPDE	594		
DB	993	CGVGQRQSVRC-----TSHGTQASHECTEALRPPTTQOCEA-KCDSPTPGDGPEE	1042		
RESULT 4					
AT53_HUMAN	AT53_HUMAN	STANDARD;	PRT; 1201 AA.		
AC	OL5072;				
DT	16-OCT-2001 (Rel. 40, Created)				
DT	16-OCT-2001 (Rel. 40, Last sequence update)				
DE	ADAMTS-3 precursor (RC 3.4.24.) (A disintegrin and metalloproteinase with thrombospondin motifs 3) (ADAM-TS 3) (ADAM-TS3) (fragment).				
GN	ADAMTS3 OR KIAA0366.				
OS	Homo sapiens (human).				
OC	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;				
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.				
OX	NCBI_TaxID:9606;				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RC	TISSUE=Brain;				
EX	MEDLINE=97349984; PubMed=9205841;				
RA	Nagase T., Ishikawa K.-I., Nakajima D., Ohira M., Seki N., Miyajima N., Tanaka A., Kotani H., Nomura N., Ohara O.;				

Prediction of the coding sequences of unidentified human genes. VII. The complete sequences of 100 new cDNA clones from brain which can code for large proteins in vitro.;
DNA Res. 4:141-150(1997).
CC - COFACTOR: BINDS 1 ZINC ION (BY SIMILARITY).
CC - SUBCELLULAR LOCATION: SECRETED. ASSOCIATED WITH THE EXTRACELLULAR MATRIX (BY SIMILARITY).
CC - DOMAIN: THE SPACER DOMAIN AND THE TSP TYPE-1 DOMAINS ARE IMPORTANT FOR A TIGHT INTERACTION WITH THE EXTRACELLULAR MATRIX.
CC - PTM: THE PRECURSOR IS CLEAVED BY A FUIN ENDOPEPTIDASE (BY SIMILARITY).
CC - SIMILARITY: BELONGS TO PEPTIDASE FAMILY M12b.
CC - SIMILARITY: CONTAINS 1 DISINTEGRIN-LIKE DOMAIN.
CC - SIMILARITY: CONTAINS 4 TSP TYPE-1 DOMAINS.
CC - CAUTION: HAS SOMETIMES BEEN REFERRED TO ADAMTS4.

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See [http://www.isb-sib.ch/](http://www.isb-sib.ch/announce/) or send an email to license@isb-sib.ch).

EMBL; AB002364; BAA20821.1; -
MEROPS; M12.220; -
MIM; 605011; -
InterPro; IPR001762; Disintegrin.
InterPro; IPR002870; Pep_M12B_propep.
InterPro; IPR001590; Reprolysin.
InterPro; IPR000884; TSP1.
InterPro; IPR000130; Zn_MTpeptidse.
Pfam; PF01562; Pep_M12B_propep; 1.
Pfam; PF01421; Reprolysin; 1.
Pfam; PF00090; tsp_1; 4.
SMART; SM00209; TSP1; 4.
PROSITE; PS0215; ADAM_MEPRO; 1.
PROSITE; PS0092; TSP1; 2.
PROSITE; PS0143; ZINC_PROTEASE; FALSE_NEG.
PROSITE; PS0042; DISINTEGRIN_1; FALSE_NEG.
KW Hydrolase; Metalloprotease; Zinc; Signal; Glycoprotein; Zymogen;
Repeat; Extracellular matrix; Heparin-binding.
FT SIGNAL <1 1 POTENTIAL.
FT PROPEP 2 245 BY SIMILARITY.
FT CHAIN 246 1201 ADAMTS-3.
FT METAL 394 394 ZINC (CATALYTIC) (BY SIMILARITY).
FT ACT_SITE 395 395 BY SIMILARITY.
FT METAL 398 398 ZINC (CATALYTIC) (BY SIMILARITY).
FT METAL 404 404 ZINC (CATALYTIC) (BY SIMILARITY).
FT DOMAIN 466 546 DISINTEGRIN-LIKE.
FT DOMAIN 547 603 TSP TYPE-1 1.
FT DOMAIN 604 708 CYS-RICH.
FT DOMAIN 709 840 SPACER.
FT DOMAIN 841 898 TSP TYPE-1 2.
FT DOMAIN 899 961 TSP TYPE-1 3.
FT DOMAIN 962 1013 TSP TYPE-1 4.
FT DOMAIN 242 245 POLY-ARG.
FT CARBOHYD 79 79 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 115 115 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 238 238 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 341 341 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 471 471 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 810 810 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 938 938 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 1201 AA; 135113 MW; D54EA92BD506A3AA CRC64;

Query Match 6.8%; Score 651; DB 1; Length 1201;
Best Local Similarity 31.3%; Pred. No. 2e-32;
Matches 181; Conservative 65; Mismatches 216; Indels 116; Gaps 27;
QY 28 SEEDRDGLDAMGPWSECSRTCGGASYSILRRCLS-----SKSCGRNRYRTCSNVDC 81

```

db 542 NOQKODGNWGWTKGSCSCSTGTVRFTRCNPNPMPINGGDCPGVNFYQLCNTTEEC 601
      ::: :| | : | | | | | : | : | : | | : | : | : | : | : | : |
QY 82 PPEAGDFRAQOC SAHN-----DVKHGOFYEWLPSVN--DPDNPCSLKQAKGTTILVEL 134
      | | | | | | | | | | : | | : | | | | | : | | : | | | | | : |
Db 602 QKHFDFAQOCCOORNSHEFYQNTKHH-----WLPYEPDPKRRCHLYCQSKGTGVAYM 656
      | | | | | | | | | | : | | : | | | | | : | | : | | | | | : |
QY 135 APKVLDTGTRC-YTESLDMCISGLICQLVGDHGLGVKEDNCVCVNGDSTCLRVGQY- 192
      | | | | | | | | | | : | | : | | | | | : | | : | | | | | : |
Db 657 KQLVHDGTHCSYKDPYSICVRGCVKVCDDKEIGSKNKNEDKCCVCGGDNSHCHRTVKGFTT 716
      | | | | | | | | | | : | | : | | | | | : | | : | | | | | : |
QY 193 --KSQLSATKSDTVAIPYSGSHRILVLK--GPDHLYLETKT-----LQGTKGNSLS 242
      | | | | | | | | | | : | | : | | | | | : | | : | | | | | : |
Db 717 RTRPKLGYLKMD-----IPGARHV-LIQDEASPHILAINKOATGHYILNG-KGEAKS 770
      | | | | | | | | | | : | | : | | | | | : | | : | | | | | : |
QY 243 STGTFVLVNSSVDQFKPEKDEILRMAGPLTADIVKI---RNSGSDSTVQVFIFYQ---P 296
      | | | | | | | | | | : | | : | | | | | : | | : | | | | | : |
Db 771 R--TFIDLGVWDYNIEDDTESLHTDGLDHPVILIIPOENDTRSLATKYKIIHEDSV 828
      | | | | | | | | | | : | | : | | | | | : | | : | | | | | : |
QY 297 IIR-----RWRETFPCSATCGGYQLTSAECYDLRSNRVVAQDYCHYYPEN 344
      | | | | | | | | | | : | | : | | | | | : | | : | | | | | : |
Db 829 TINSNNVIQELDTFEWALKSWQSVKPCGGGFGYTKYGERKSDNKMVHSEFC---AN 885
      | | | | | | | | | | : | | : | | | | | : | | : | | | | | : |
QY 345 IKPKPLQECNLDPCPASDGKQIMPYDLYHPLRWEATPWTACSSSCG--GGIQSRVASC 403
      | | | | | | | | | | : | | : | | | | | : | | : | | | | | : |
Db 886 KKPPIRMCMNIQECT-----HPL--WVAEEWHTKTCGSGGYQLRTVRC 929
      | | | | | | | | | | : | | : | | | | | : | | : | | | | | : |
QY 404 VEEDIQGHVTSVEWKCMYTKPKPIAOPCNIFDPC-KWLQAQWSPCHVTGCGGLRYAVVL 462
      | | | | | | | | | | : | | : | | | | | : | | : | | | | | : |
Db 930 LQPLLDGTRNSVHSKYCM-GDRPESRRPCNVPQPAQWKTGPWSECVSTGCGTEFVQVL 988
      | | | | | | | | | | : | | : | | | | | : | | : | | | | | : |
QY 463 CIDHRGMHTGCSKTKPHIKECIVP---TPCYKPK-----KLPVEAKL- 505
      | | | | | | | | | | : | | : | | | | | : | | : | | | | | : |
Db 989 C--RAGDHCDG-----EKPSVRAQLPQNDPFCGLGDKSIFCQMEVLARYCSIPGYNKLC 1042
      | | | | | | | | | | : | | : | | | | | : | | : | | | | | : |
QY 506 -----PWFKQAQELERGAEEPSPIFPA 530
      | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 1043 CESCKSRSTLPPPYLLAEATHDD--VISNPSDLPRS 1078
      | : | : | : | : | : | : | : | : | : | : | : | : | : | : |

RESULT 5
ATS2 BOVIN STANDARD; PRT; 1205 AA.
AC AC P79331;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE ADAMTS-2 precursor (EC 3.4.24.14) (A disintegrin and
DE metalloproteinase with thrombospondin motifs 2) (ADAM-TS2)
DE (Procollagen I/II amino-propeptide processing enzyme) (Procollagen I
DE N-proteinase) (PC I-NP) (Procollagen N-endopeptidase) (pNPI).
DE ADAMTS2 OR NPI.
GN Bos taurus (Bovine).
OS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OC Bovidae; Bovinae; Bos.
OX NCBI_TaxID=9913;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Skin;
RX MEDLINE=97225960; PubMed=9122202;
RA Collige A., Li S.W., Sieron A.L., Nusgens B.V., Prockop D.J.,
RA Lapiere C.M.;
RT "cDNA cloning and expression of bovine procollagen I N-proteinase: a
RT new member of the superfamily of zinc-metalloproteinases with binding
RT sites for cells and other matrix components.";
RL Proc. Natl. Acad. Sci. U.S.A. 94:2374-2379(1997).
RN [2]
RP PARTIAL SEQUENCE.
RX MEDLINE=95348096; PubMed=7622483;
RA Collige A., Beschlin A., Samyn B., Goebels Y., Van Beeumen J.,
RA Nusgens B.V., Lapiere C.M.;
RT "Characterization and partial amino acid sequencing of a 107-kDa

```

```

RT procollagen I N-proteinase purified by affinity chromatography on
RT immobilized type XIV collagen.";
RL J. Biol. Chem. 270:16724-16730(1995).
CC -!- FUNCTION: CLEAVES THE PROPEPTIDES OF TYPE I AND II COLLAGEN PRIOR
CC TO FIBRIL ASSEMBLY. DOES NOT ACT ON TYPE III COLLAGEN. MAY ALSO
CC PLAY A ROLE IN DEVELOPMENT THAT IS INDEPENDANT OF ITS ROLE IN
CC COLLAGEN BIOSYNTHESIS.
CC -!- CATALYTIC ACTIVITY: Cleaves the N-propeptide of collagen chain
CC alpha-1(I) at Pro-|-Gln and of alpha-1(II) and alpha-2(I) chains
CC at Ala-|-Gln.
CC -!- COFACTOR: BINDS 1 ZINC ION (BY SIMILARITY).
CC -!- SUBUNIT: MAY BELONG TO A MULTIMERIC COMPLEX. BINDS SPECIFICALLY TO
CC COLLAGEN TYPE XIV.
CC -!- SUBCELLULAR LOCATION: SECRETED. ASSOCIATED WITH THE EXTRACELLULAR
CC MATRIX (BY SIMILARITY).
CC -!- TISSUE SPECIFICITY: ENZYMIC ACTIVITY IS DETECTED AT HIGH LEVEL
CC IN ALL TYPE I COLLAGEN-RICH TISSUES SUCH AS SKIN, BONES, TENDONS
CC AND AORTA AND AT LOW LEVEL IN BRAIN AND THYMUS. THE MRNA LEVELS
CC WERE DISPROPORTIONATELY HIGH IN HEART, LIVER, RETINA AND MUSCLE.
CC -!- DOMAIN: THE SPACER DOMAIN AND THE TSP TYPE-1 DOMAINS ARE IMPORTANT
CC FOR A TIGHT INTERACTION WITH THE EXTRACELLULAR MATRIX.
CC -!- PTM: THE N-TERMINUS IS BLOCKED.
CC -!- PTM: THE PRECURSOR IS CLEAVED BY A FURIN ENDOPEPTIDASE (BY
CC SIMILARITY).
CC -!- DISEASE: DEFECTS IN ADAMTS2 ARE THE CAUSE OF DERMATOSPARAXIS, A
CC RECESSIVELY INHERITED DISORDER CHARACTERIZED BY SEVERE SKIN
CC FRAGILITY AND BIOCHEMICALLY BY THE PRESENCE IN SKIN OF PROCOLLAGEN
CC INCOMPLETELY PROCESSED AT THE AMINO TERMINUS.
CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M12B.
CC -!- SIMILARITY: CONTAINS 1 DISINTEGRIN-LIKE DOMAIN.
CC -!- SIMILARITY: CONTAINS 4 TSP TYPE-1 DOMAINS.
CC -!- CAUTION: HAS SOMETIMES BEEN REFERRED TO ADAMTS3.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: X96389; CAA65253.1; -.
DR InterPro: IPR001762; Disintegrin.
DR InterPro: IPR002870; Pep_M12B_propep.
DR InterPro: IPR001590; Reprolysin.
DR InterPro: IPR000884; TSP1.
DR InterPro: IPR000130; Zn_Mtpeptidse.
DR Pfam: PF01562; Pep_M12B_propep; 1.
DR Pfam: PF01421; Reprolysin; 1.
DR Pfam: PF00090; tsp_1; 4.
DR SMART: SM00209; TSP1; 4.
DR PROSITE: PS50215; ADAM_MPEPRO; 1.
DR PROSITE: PS50092; TSP1; 1.
DR PROSITE: PS00142; ZINC_PROTEASE; 1.
DR PROSITE: PS00427; DISINTEGRIN_1; FALSE NEG.
KW Hydrolase; Metalloproteinase; Zinc; Signal; Glycoprotein; Zymogen;
KW Repeat; Collagen degradation; Extracellular matrix; Heparin-binding.
FT SIGNAL 1 28
FT PROPEP 29 253
FT CHAIN 254 1205
FT METAL 402 403
FT ACT_SITE 403 403
FT METAL 406 406
FT METAL 412 412
FT DOMAIN 474 554
FT DOMAIN 555 611
FT DOMAIN 612 716
FT DOMAIN 717 845
FT DOMAIN 846 905
FT DOMAIN 906 968
FT DOMAIN 969 1024
FT SITE 685 687
FT DOMAIN 31 35

```


[illegible]

DR InterPro; IPR000884; TSP1.
 DR InterPro; IPR000130; Zn_Mtpeptdse.
 DR Pfam; PF01562; Pep_M12B_propep; 1.
 DR Pfam; PF01421; Repolysin; 1.
 DR Pfam; PF00090; TSP_1; 3.
 DR SMART; SM00209; TSP1; 3.
 DR PROSITE; PS0215; ADAM_MEPRO; 1.
 DR PROSITE; PS00142; ZINC_PROTEASE; 1.
 DR PROSITE; PS00427; DISINTEGRIN_1; FALSE NEG.
 KW Hydrolase; Metalloprotease; Zinc; Signal; Glycoprotein; Zymogen;
 KW Repeat; Extracellular matrix; Heparin-binding.
 FT SIGNAL 1 54
 FT PROPEP 55 252
 FT CHAIN 253 967
 FT SITE 205 205
 FT METAL 401 401
 FT ACT_SITE 402 402
 FT METAL 405 405
 FT METAL 411 411
 FT DOMAIN 476 518
 FT DOMAIN 559 616
 FT DOMAIN 616 724
 FT DOMAIN 725 857
 FT DOMAIN 858 907
 FT DOMAIN 908 967
 FT DOMAIN 194 198
 FT CARBOHYD 547 547
 FT CARBOHYD 720 720
 FT CARBOHYD 764 764
 FT CARBOHYD 782 782
 FT CARBOHYD 945 945
 FT CONFLICT 21 21
 FT CONFLICT 26 31
 FT CONFLICT 49 49
 FT CONFLICT 72 72
 FT CONFLICT 79 79
 FT CONFLICT 249 249
 FT CONFLICT 262 265
 FT CONFLICT 607 607
 FT CONFLICT 936 936
 FT CONFLICT 962 962
 SQ SEQUENCE 967 AA; 105705 MW; F93C864F6DCDB4CF CRC64;
 Query Match 6.4%; Score 613; DB 1; Length 967;
 Best Local Similarity 29.6%; Pred. No. 3.3e-30;
 Matches 145; Conservative 65; Mismatches 166; Indels 114; Gaps 19;
 QY 34 GLWDAMPWSECSRTCGGASYSLRRC-----LSSKSCGRNTRYRTCSNVDCPPAG- 86
 DB 560 GSWGPWGPWDCSNTCGGVQYTMRECDNVPKNGKGYCEGKRYRYSNCTEDCPDNNGK 619
 QY 87 DFRQAQCSAHNDYKH-----HQGFVWLP--VSNPDNPSCLKCAKGTLLVVELAPKVL 140
 DB 620 TFREOQCEAAHNEFSAKSGNEPTWTPKYAGVPKDKCKTCEAKGIGYFVLQPKVVD 679
 QY 141 GTRCYTESLDMCISGLCQIVGCDHQIGSTYKEDNCGVNGDSTCRIVRGQYKSQLSATK 200
 DB 680 GTPCSPDSTVCVQGVKACDRIDSKKKFKDCGVCGNGSTCKKISG---FVTSIRP 736
 QY 201 SDDTVAIPYSGRIHLVKGPDHLYLETKLQTKGENSL-----SSTGTFLDVN----S 252
 DB 737 GYHDVITPAGATNIEVKHRNP-----RGRNNGSFALRAADGTYLINGNFTLS 786
 QY 253 SVDFQKFDKELMAGPLTADFTVKIRNSGADS--TVQFIYQPIIHRWRETDFPCCS 310
 DB 787 TLEQDLTKGIVLRVSGSSAA--LIRNFSFLKEPLTIQVLMV----- 828
 QY 311 ATCGGGYQLTSAECYDLRSNRWADQVCHYYPENIKPKIQLQCNLDPCPASDGY----- 365
 DB 829 -----GHAL-----RPKIK-----YTFPMK 843

QY 366 KQIMPYDLYHPLPRWEATPWTACSSSGGSIOSRAVSCVEEDIQGHVTSVEENKCMVTPK 425
 DB 844 KKTEPEFNALPTTFSEWVIEWEGSCSKTCGGQWQRVVEEC--RDINGPAS-----EKAKEVK 897
 QY 426 MPFAQPCNLFIDCPKWLQAEWSPCTVTCGGRLRYRVVLICIDHRG--MHTGCSPTKTP-HI 482
 DB 898 PASTRCADLPFCPRWQVGDWSPCKTCGKGYKRTKLCLSHDGGVLSNESCDELKPKPKHY 957
 QY 483 KEECIVPTTPC 492
 DB 958 IDFCIL-TQC 966
 RESULT 7
 AT2S2_HUMAN
 ID AT2S2_HUMAN STANDARD; PRT; 1211 AA.
 AC O95450;
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE ADAMTS-2 precursor (EC 3.4.24.14) (A disintegrin and
 DE metalloproteinase with thrombospondin motifs 2) (ADAM-TS 2) (ADAM-TS2)
 DE (Procollagen I/II amino-propeptide processing enzyme) (Procollagen I
 DE N-proteinase) (PC I-NP) (Procollagen N-endopeptidase) (pNPI)
 DE (Procollagen I/II amino-propeptide processing enzyme).
 GN ADAMTS2 OR PCNP1 OR PCPNI.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A. (ISOFORMS LPNPI AND SPNPI).
 RC TISSUE=Skin;
 RX MEDLINE=99347935; PubMed=10417273;
 RA Collige A., Sieron A.L., Li S.-W., Schwarze U., Petty E.,
 RA Wiertelacki W., Wilcox W., Krakow D., Cohn D.H., Reardon W.,
 RA Byers P.H., Lapiere C.M., Prockop D.J., Nusgens B.V.;
 RT "Human Ehlers-Danlos syndrome type VII C and bovine dermatosparaxis
 RT are caused by mutations in the procollagen I N-proteinase gene.";
 RL Am. J. Hum. Genet. 65:308-317(1999).
 CC -!- FUNCTION: CLEAVES THE PROPEPTIDES OF TYPE I AND II COLLAGEN PRIOR
 CC TO FIBRIL ASSEMBLY. DOES NOT ACT ON TYPE III COLLAGEN. MAY ALSO
 CC PLAY A ROLE IN DEVELOPMENT THAT IS INDEPENDANT OF ITS ROLE IN
 CC COLLAGEN BIOSYNTHESIS.
 CC -!- CATALYTIC ACTIVITY: Cleaves the N-propeptide of collagen chain
 CC alpha-1(I) at Pro-I-Gln and of alpha-1(II) and alpha-2(I) chains
 CC at Ala-I-Gln.
 CC -!- COFACTOR: BINDS 1 ZINC ION (BY SIMILARITY).
 CC -!- SUBUNIT: MAY BELONG TO A MULTIMERIC COMPLEX. BINDS SPECIFICALLY TO
 CC COLLAGEN TYPE XIV (BY SIMILARITY).
 CC -!- SUBCELLULAR LOCATION: SECRETED. ASSOCIATED WITH THE EXTRACELLULAR
 CC MATRIX (BY SIMILARITY).
 CC -!- ALTERNATIVE PRODUCTS: 2 ISOFORMS; LPNPI (SHOWN HERE) AND SPNPI;
 CC ARE PRODUCED BY ALTERNATIVE SPLICING. SPNPI HAS NO SIGNIFICANT N-
 CC PROCOLLAGEN PEPTIDASE ACTIVITY.
 CC -!- TISSUE SPECIFICITY: EXPRESSED AT HIGH LEVEL IN SKIN, BONE, TENDON
 CC AND AORTA AND AT LOW LEVELS IN THYMUS AND BRAIN.
 CC -!- DOMAIN: THE SPACER DOMAIN AND THE TSP TYPE-1 DOMAINS ARE IMPORTANT
 CC FOR A TIGHT INTERACTION WITH THE EXTRACELLULAR MATRIX.
 CC -!- PTM: THE PRECURSOR IS CLEAVED BY A FURIN ENDOPEPTIDASE (BY
 CC SIMILARITY).
 CC -!- DISEASE: DEFECTS IN ADAMTS2 ARE THE CAUSE OF EHLERS-DANLOS
 CC SYNDROME TYPE VIIC (EDS-VIIC). A RECESSIVELY INHERITED DISORDER
 CC CHARACTERIZED CLINICALLY BY SEVERE SKIN FRAGILITY AND
 CC BIOCHEMICALLY BY THE PRESENCE IN SKIN OF PROCOLLAGEN INCOMPLETELY
 CC PROCESSED AT THE AMINO TERMINUS.
 CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M12B.
 CC -!- SIMILARITY: CONTAINS 1 DISINTEGRIN-LIKE DOMAIN.
 CC -!- SIMILARITY: CONTAINS 4 TSP TYPE-1 DOMAINS.
 CC -!- CAUTION: HAS SOMETIMES BEEN REFERRED TO ADAMTS3.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

QY	198	ATKSD	-----TVVAIPYGSRRHRLVRLKPDHLYLETKTLOGK-----GENSL-SSTGTF	247
Db	730	-TRSPKKGHYIKMFEIPAGARHLLIQEVDATSHHLLAVKNLETKGKILNEENDVDASSKTF	788	
QY	248	LVDNSSVDFOKFPPOKEILLRMAGPLTADTIVKIRNCGSADSTVOFTYOPITIHWRWETDFF	307	
Db	789	IAMGVWEIREDDEGRELQTMGLPGLHTITVLIVIPVG--DTRVS-LTYKYMIH-----	838	
QY	308	PCSAICGGGYOLTSAECYDLRSNRVADQYCHYYPENIKPKPKLQECNLDPCPASDGYKQ	367	
Db	839	-----DSLVNDNNVLEEDSVY-----	856	
QY	368	IMPYDLHPPLPWEATFTWACSSCGGQISRAVSCVEEDIQGHVTSVEEWKCMYTPKMP	427	
Db	857	-----EWALKWSPCKPCGGGSGQTKYGC-----	881	
QY	428	IAOPCNIFDCPKWLAQEWSPCTVTCGGRLRYRVVLCIDHRGMHTGGCSPKTKPHKEPCI	487	
Db	882	-----RRR-----LDHKMVHRGFCALSKPKAIRAC	908	
QY	488	VPTPCYKPKELPVEAKLPWFKAQLEGAASVEEPSFIPKAWSACTVTCG-VGTQVRI	546	
Db	909	NPOEC-----SQPVWTGWEFCSQTCGRGTGMQVRS	939	
QY	547	VRCQVLLSFQSADLPIDCEGPKPASQACVAGPCSGEIPFNPDETDLGLFGGLQDFD	606	
Db	940	VRC-----IQPLHD-----NTTRSVHAKHCNDARPE-----	965	
QY	607	EHYDWEYEGFTKCSGCGGVQEAUVCLNKQTPREAPENLCVTSRPPQLLKSCNLDPC	666	
Db	966	-----SR-----ACSRCL	975	
QY	667	PARWETKGSPOSLTCGVGLQTRDVFCSHLLSREMNETVILADE--LCROPKPSVQAC	723	
Db	976	PCRWRAGPSQCSVTGCTGTQRPVPCR-----TADDSFGICQDEEPTARTC	1023	
QY	724	NRFNCP-----PAWPAQW-----QPCSR-----CGGVQKREVLCKORMADGSF	764	
Db	1024	RLGPCRNLSDESKSYVQVLSRPPDPSPIRKSSKGHCQG--DKSIFCR-----	1072	
QY	765	LELPETFCSA--SKPACQ	780	
Db	1073	MEVLSRYCSIPGYNKLCK	1091	
RESULT	8			
ATSL_MOUSE				
ID	ATSL_MOUSE	STANDARD;	PRT;	968 AA.
AC	P97857; O54768;			
DT	30-MAY-2000 (Rel. 39, Created)			
DT	16-OCT-2001 (Rel. 40, Last sequence update)			
DE	16-OCT-2001 (Rel. 40, Last annotation update)			
DE	ADAMTS-1 precursor (EC 3.4.24.-) (A disintegrin and metalloproteinase with thrombospondin motifs 1) (ADAM-TS 1) (ADAM-TS1).			
GN	ADAMTS1			
OS	Mus musculus (Mouse).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.			
OX	NCBI_TaxID=10090;			
RP	[1]			
RN	SEQUENCE FROM N.A.			
RC	STRAIN=129/SVJ;			
RC	MEDLINE=98110583; PubMed=9441751;			
RX	Kuno K., Lizasa H., Ohno S., Matsushima K.;			
RT	"The exon/intron organization and chromosomal mapping of the mouse			
RT	ADAMTS-1 gene encoding an ADAM family protein with TSP motifs.";			
RL	Genomics 46:466-471(1997).			
RN	[2]			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE=97150761; PubMed=8995297;			
RA	Kuno K., Kanada N., Nakashima E., Fujiki F., Ichimura F.;			
RA	Matsushima K.;			
RT	"Molecular cloning of a gene encoding a new type of metalloproteinase-			

the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announcement/> or send an email to license@isb-sib.ch).

EMBL: AJ003125; CRA05880.1; -.
MEROPS; M12.301; -.
MIM; 604539; -.
MIM; 225410; -.
InterPro; IPR001762; Disintegrin.
InterPro; IPR002870; Pep_M12B_propep.
InterPro; IPR001590; Reprolysin.
InterPro; IPR000884; TSPI.
InterPro; IPR000130; Zn_MTpeptdse.
Pfam; PF01562; Pep_M12B_propep; 1.
Pfam; PF01421; Reprolysin; 1.
Pfam; PF00090; tsp_1; 4.
SMART; SM00209; TSPI; 4.
PROSITE; PS0215; ADAM_MEPRO; 1.
PROSITE; PS50092; TSPI; 1.
PROSITE; PS00142; ZINC_PROTEASE; FALSE_NEG.
PROSITE; PS00427; DISINTEGRIN_1; FALSE_NEG.
Hydrolase; Metalloprotease; Zinc; Signal; Glycoprotein; Zymogen;
Repeat; Collagen degradation; Extracellular matrix; Heparin-binding;
Alternative splicing;
KW SIGNAL; 1 29 POTENTIAL.
KW PROPEP; 30 253 BY SIMILARITY.
FT CHAIN; 254 1211 ADAMTS-2.
FT METAL; 408 408 ZINC (CATALYTIC) (BY SIMILARITY).
FT ACT_SITE; 409 409 BY SIMILARITY.
FT METAL; 412 412 ZINC (CATALYTIC) (BY SIMILARITY).
FT METAL; 418 418 ZINC (CATALYTIC) (BY SIMILARITY).
FT DOMAIN; 480 560 ZINC (CATALYTIC) (BY SIMILARITY).
FT DOMAIN; 561 617 TSP TYPE-1 1.
FT DOMAIN; 618 722 CYS-RICH.
FT SITE; 691 693 CELL ATTACHMENT SITE (POTENTIAL).
FT DOMAIN; 723 851 SPACER.
FT DOMAIN; 852 911 TSP TYPE-1 2.
FT DOMAIN; 912 974 TSP TYPE-1 1 3.
FT DOMAIN; 975 1030 TSP TYPE-1 4.
FT DOMAIN; 40 43 POLY-ALA.
FT DOMAIN; 185 188 POLY-GLU.
FT CARBOHYD; 112 112 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD; 251 251 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD; 949 949 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD; 993 993 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD; 1031 1031 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD; 1098 1098 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD; 1145 1145 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD; 1150 1150 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT VARSPLIC; 544 566 HCPEGLCWLPDLKRDGWSGA -> FRGVAHAACYP
FT VARSPLIC; 567 1211 TLGGQHWIA (IN ISOFORM SPNP1).
FT SEQUENCE; 1211 AA; 134722 MW; BECEE25C23CAD2D CRC64;
MISSING (IN ISOFORM SPNP1).
Query Match 6.3%; Score 608.5; DB 1; Length 1211;
Best Local Similarity 22.9%; Pred. No. 8.6e-30;
Matches 183; Conservative 87; Mismatches 212; Indels 317; Gaps 29;
32 RDGLWDANGPWSRCSRTCGGGASYSLRRC-----LSSKSCGRNIRYRCSNVDCPPEA 85
560 RDGSGWGAAPFGSCSRCTCGTGVKFRHQCDNPHANGRTCSGLAYDFQLCSRODCPDSL 619
86 GDFRAQOCS-----AINDKVHHGQFVEWLPVSN-DPDNPSCLKCAQKTTLVVELAPKV 138
620 ADFEEQCRQWDLYFEHGDQHH-----WLPHEHRAKERCHLYCESRESGETGVVSMKRV 674
139 LDGTRC-YTESLDMCISGLCOIVGCDHOLGTSKVDKNGCVCNGDGTCTCLRVGOYKSQLS 197
675 HDGTRCSYKDAFSLCYRGDCRKVCGDGVIGSSQKQDKCGCGGDNHCKVYKGTf--- 729

AC Q9UR18; Q9UP80; Q9UH83; Q9P2K0; Q9NSJ8;
 DT 30-MAY-2000 (Rel. 39, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 01-MAR-2002 (Rel. 41, Last annotation update)
 DE ADAMTS-1 precursor (EC 3.4.24.-) (A disintegrin and metalloproteinase
 DE with thrombospondin motifs 1) (ADAM-TS 1) (ADAM-TS1) (METH-1).
 GN ADAMTS1 OR METH1 OR KIAA1346.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RN SEQUENCE FROM N.A.
 RA Casas C., Pritchard M.A., Estivill X., Arbones M.L.;
 RA "Cloning, characterization and mapping on human chromosome 21 of the
 RA orthologue of murine Adamts-1.";
 RA Submitted (JUL-1999) to the EMBL/GenBank/DBJ databases.
 RL [2]
 RL SEQUENCE FROM N.A., AND FUNCTION.
 RP TISSUE=Heart;
 RC MEDLINE=99367466; PubMed=10438512;
 RA Vazquez F., Hastings G., Ortega M.-A., Lane T.F., Oikemus S.,
 RA Lombardo M., Iruela-Arispe M.L.;
 RA "METH-1, a human ortholog of ADAMTS-1, and METH-2 are members of a new
 RA family of proteins with angio-inhibitory activity.";
 RL J. Biol. Chem. 274:23349-23357(1999).
 RN [3]
 RN SEQUENCE FROM N.A.
 RP TISSUE=Endothelial cells;
 RC MEDLINE=20247184; PubMed=10785405;
 RA Glienke J., Schmitt A.O., Pilarsky C., Hinzmann B., Weiss B.,
 RA Rosenthal A., Thierach K.H.;
 RA "Differential gene expression by endothelial cells in distinct
 RA angiogenic states";
 RL Eur. J. Biochem. 267:2820-2830(2000).
 RN [4]
 RN SEQUENCE FROM N.A.
 RP TISSUE=Brain;
 RC MEDLINE=20181126; PubMed=10718198;
 RA Nagase T., Kikuno R., Ishikawa K.-I., Hirotsawa M., Ohara O.;
 RA "Prediction of the coding sequences of unidentified human genes.XVI.
 RA The complete sequences of 150 new cDNA clones from brain which code
 RA for large proteins in vitro.";
 RL DNA Res. 7:65-73(2000).
 RN [5]
 RN SEQUENCE FROM N.A.
 RP MEDLINE=20289799; PubMed=10830953;
 RA Battori M., Fujiyama A., Taylor T.D., Watanabe H., Yada T.,
 RA Park H.-S., Toyoda A., Ishii K., Totoki Y., Choi D.-K., Soeda E.,
 RA Ohki M., Takagi T., Sakaki Y., Taudien S., Blechschmidt K., Polley A.,
 RA Menzel U., Delabar J., Kumpf K., Lehmann R., Patterson D.,
 RA Reichwald K., Rump A., Schillhabel M., Schudy A., Zimmermann W.,
 RA Rosenthal A., Kudoh J., Shibuya K., Kawasaki K., Asakawa S.,
 RA Shintani A., Sasaki T., Nagamine K., Mitsuyama S., Antonarakis S.E.,
 RA Minoshima S., Shimizu N., Nordsiek G., Hornischer K., Brandt P.,
 RA Scharfe M., Schoen O., Desario A., Reichelt J., Kauer G., Bloeker H.,
 RA Wehrmeyer S., Borzym K., Gardiner K., Hennig S., Kieselmann L., Dagan E.,
 RA Leirach H., Reinhardt R., Raspo M.-L.;
 RA "The DNA sequence of human chromosome 21.";
 RL Nature 405:311-319(2000).
 RN [6]
 RN SEQUENCE OF 418-967 FROM N.A.
 RP TISSUE=Melanoma;
 RC Blum H., Bauersachs S., Mewes H.-W., Weil B., Wiemann S.;
 RA Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
 CC -1- FUNCTION: CLEAVES AGGREGAN, A CARTILAGE PROTEOGLYCAN, AND MAY BE
 CC INVOLVED IN ITS TURNOVER (BY SIMILARITY). HAS ANGIOGENIC INHIBITOR
 CC ACTIVITY. ACTIVE METALLOPROTEINASE, WHICH MAY BE ASSOCIATED WITH
 CC VARIOUS INFLAMMATORY PROCESSES AS WELL AS DEVELOPMENT OF CANCER
 CC CACHEXIA. MAY PLAY A CRITICAL ROLE IN FOLLICULAR RUPTURE.
 CC -1- CATALYTIC ACTIVITY: CLEAVES AGGREGAN AT THE 1938-GLU-1-LEU-1939
 CC SITE, WITHIN THE CHONDROITIN SULFATE ATTACHMENT DOMAIN.

CC -1- COFACTOR: BINDS 1 ZINC ION (BY SIMILARITY).
 CC -1- SUBCELLULAR LOCATION: SECRETED. ASSOCIATED WITH THE EXTRACELLULAR
 CC MATRIX (BY SIMILARITY).
 CC -1- DOMAIN: THE SPACER DOMAIN AND THE TSP TYPE-1 DOMAINS ARE IMPORTANT
 CC FOR A TIGHT INTERACTION WITH THE EXTRACELLULAR MATRIX.
 CC -1- PTM: THE PRECURSOR IS CLEAVED BY A URIN ENDOPEPTIDASE (BY
 CC SIMILARITY).
 CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M12B.
 CC -1- SIMILARITY: CONTAINS 1 DISINTEGRIN-LIKE DOMAIN.
 CC -1- SIMILARITY: CONTAINS 3 TSP TYPE-1 DOMAINS.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; AF170084; AAF15317.1; -
 CC EMBL; AF060152; AAD48080.1; ALT_INIT.
 CC EMBL; AF207664; AAF23772.1; -
 CC EMBL; AB037767; BAA92584.1; ALT_INIT.
 CC EMBL; AP001697; BAA95502.1; -
 CC EMBL; AL162080; CAB82413.1; -
 CC MIN; 605174; -
 CC InterPro: IPR001762; Disintegrin.
 CC InterPro: IPR002870; pep_M12B_propep.
 CC InterPro: IPR001590; Reprolysin.
 CC InterPro: IPR000884; TSP1.
 CC InterPro: IPR000130; zn_Mtpeptidse.
 CC Pfam: PF01562; Pep_M12B_propep; 1.
 CC Pfam: PF01421; Reprolysin; 1.
 CC Pfam: PF00090; tsp_1; 3.
 CC SMART; SM00209; TSP1; 3.
 CC PROSITE; PS00215; ADAM_MEPRO; 1.
 CC PROSITE; PS00427; DISINTEGRIN_1; FALSE_NEG.
 CC PROSITE; PS00092; TSP1; 2.
 CC PROSITE; PS00142; ZINC_PROTEASE; 1.
 CC Hydrolase; Metalloprotease; zinc; Signal; Glycoprotein; Zymogen;
 CC Repeat; Extracellular matrix; Heparin-binding.
 CC SIGNAL 1 49
 CC FT PROPEP 50 252 BY SIMILARITY.
 CC FT CHAIN 253 967 ADAMTS-1.
 CC FT SITE 198 198 CYSTEINE SWITCH (POTENTIAL).
 CC FT METAL 401 401 ZINC (CATALYTIC) (BY SIMILARITY).
 CC FT ACT_SITE 402 402 BY SIMILARITY.
 CC FT METAL 405 405 ZINC (CATALYTIC) (BY SIMILARITY).
 CC FT METAL 411 411 ZINC (CATALYTIC) (BY SIMILARITY).
 CC FT DOMAIN 476 559 DISINTEGRIN-LIKE.
 CC FT DOMAIN 560 616 TSP TYPE-1 1.
 CC FT DOMAIN 617 724 CYS-RICH.
 CC FT DOMAIN 725 849 SPACER.
 CC FT DOMAIN 850 908 TSP TYPE-1 2.
 CC FT DOMAIN 909 967 TSP TYPE-1 3.
 CC FT DOMAIN 843 846 POLY-LYS.
 CC FT CARBOHYD 547 547 N-LINKED (GLCNAC. . .) (POTENTIAL).
 CC FT CARBOHYD 720 720 N-LINKED (GLCNAC. . .) (POTENTIAL).
 CC FT CARBOHYD 764 764 P -> A (IN REF. 4 AND 5).
 CC FT CONFLICT 227 227 CONFLICT (GLCNAC. . .) (POTENTIAL).
 CC FT CONFLICT 468 468 O -> H (IN REF. 1).
 CC FT CONFLICT 561 561 S -> N (IN REF. 1).
 CC FT SEQUENCE 967 AA; 105383 MW; C189389324741EDI CRC64;

Query Match 6.1%; Score 584.5; DB 1; Length 967;
 Best Local Similarity 29.4%; Pred. No 1.9e-28;
 Matches 148; Conservative 59; Mismatches 168; Indels 129; Gaps 19;
 QY 23 SRTARSEEDR--DGLMDANGPWSCESTRCTGGGASYSLLRRC-----LSSKSCSGRNIRYR 74
 DB 547 NKIDRRKHEDTPFHGSGMGMGPGDCSDCTCGGGVQYTMRECDNPVKNKGKCYCEGKRVYR 506

QY 75 TCSNVDCEPEAG--DRAQCSAHNDVKH-----HGQFYELP--VSNDDPNPQSLKCOAGK 127
Db 607 SCNLEDCPDNNKGTFRFEEQCEAHNEFSKASFGSPAVEMLPKYAGVSPKDRCKLIQCOAG 666
QY 128 TTLVVELAPKVLDTGTRCYTESLDMCISGLCOIVGCDHOLGISTVKEDNGVCGNGDSTCR 187
Db 667 IGVFFVLQPKVVDGTPCSPDSTSVCGQGVKAGCDRIIDSKKFKDKCGVCGGNGSTCKK 726
QY 188 VRGOYKQSLSAVKSDDTVAIPGSGHRLVRLKGPDLHLYETKTLQGTGKENSLSL---SS 243
Db 727 ISG---SVTSAPKPGYHDIITIPFGARINE-----VKQRNGSRNNGSFLAIKAA 773
QY 244 TGTFLVNDSSVDFOKFPDKELRMAGPLATADFIKIRNSGSDSTV-QTIFYQPIIHRWR 302
Db 774 DGYI-----LNGDYTL-----STLEQDLMYKGVVLR-- 800
QY 303 ETDFPPCSATCGGGYQLTSAECYDLNSNRVAD-----QYCHYYPENI 345
Db 801 -----YSGSSAALERIRSFPLKEPLTIOVLTGNALRPKIKYIFV--- 842
QY 346 KPKPKLQECNLDPASDGVKQIMPYDLHPLRWEATPWTACSSCGGGTQSRVAVSVE 405
Db 843 --KKKESFNAIP-----TFSAWVIEEWCECSKSCELGWQRLVEC-- 881
QY 406 EDIQGHVTSVEEWKCMYTPKMPIAOPCNIFDCPKWLAQWSPCTVTTCGGLRYRVVLCID 465
Db 882 RDINGQPAS---ECAKEVPASTRCPADHPCQWOLGESSCKTKCGKYKKRSLKCLS 937
QY 466 HRG--MHTGCSPTKTP-HIKREC 486
Db 938 HDGGVLSHESCDPLKKPKHPIDFC 961

RESULT 10
AT10_MOUSE STANDARD; PRT; 450 AA.
ID P58459;
DT 01-MAR-2002 (Rel. 41, Created)
DT 01-MAR-2002 (Rel. 41, Last sequence update)
DT 01-MAR-2002 (Rel. 41, Last annotation update)
DE ADAMTS-10 (EC 3.4.24.-) (A disintegrin and metalloproteinase with
thrombospondin motifs 10) (ADAM-TS10) (Fragment).
GN ADAMTS10.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
[1]
RN SEQUENCE FROM N.A.
RA Seldin M.F., Apte S.S.;
RT "A novel gene of the ADAMTS family predicts ADAMTS-10, a
metalloproteinase with unique structural features and expression
pattern";
RL Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.
CC -1- COFACTOR: BINDS 1 ZINC ION (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: SECRETED. ASSOCIATED WITH THE EXTRACELLULAR
MATRIX (BY SIMILARITY).
CC -1- DOMAIN: THE SPACER DOMAIN AND THE TSP TYPE-1 DOMAINS ARE IMPORTANT
FOR A TIGHT INTERACTION WITH THE EXTRACELLULAR MATRIX (BY
similarity).
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M12B.
CC -1- SIMILARITY: CONTAINS AT LEAST 4 TSP TYPE-1 DOMAINS.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (see [http://www.isb-sib.ch/](http://www.isb-sib.ch/announce/)
or send an email to license@isb-sib.ch).
CC
CC EMBL; AF302012; AAK97226.1; -
DR PROSITE; P550092; TSP1; 1.
DR

KW Hydrolase; Metalloprotease; Zinc; Glycoprotein; Repeat;
KW Extracellular matrix.
FT NON_TER 1
FT DOMAIN <1 51 CYS-RICH.
FT DOMAIN 52 174 SPACER.
FT DOMAIN 171 232 TSP TYPE-1 1.
FT DOMAIN 234 290 TSP TYPE-1 2.
FT DOMAIN 294 348 TSP TYPE-1 3.
FT DOMAIN 353 399 TSP TYPE-1 4.
FT CARBOHYD 86 86 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 141 141 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 238 238 N-LINKED (GLCNAC. .) (POTENTIAL).
SQ SEQUENCE 450 AA; 48861 MW; 2377D0E2CFBFC6A CRC64;

Query Match 5.9%; Score 564.5; DB 1; Length 450;
Best Local Similarity 30.9%; Pred. No. 1.2e-27;
Matches 151; Conservative 60; Mismatches 176; Indels 101; Gaps 21;

QY 135 APKVLDTGTRCYTESLDMCISGLCOIVGCDHOLGISTVKEDNGVCGNGDSTCRVGRGOYKS 194
Db 1 AAAYVDGTPCRPDIVDICVSGECKHVGCDRVLGDLRDKCRVCGGDSACETIEGVSP 60
QY 195 QLSATKSDDTVAIPYGSRIH-----RLVLKGPDLHLYETKTLQGTGKENSLSST 244
Db 61 ALPGTGIED-VVWIPKGSVHIFIQDLNLSLHLKLG-DOESLLEGLPGTPQPHRLPLA 118
QY 245 GTFLVNDSSVDFOKFPDK-EILRMAGPLATADFIKIRNSGSDSTVQIFYOPII----- 298
Db 119 GT-----TFHLRQGPDAQSLEALGPINASLIIMVLAQALPA-LHYRFNAPIARDALP 171
QY 299 -HRWRETDFFPCSATCGGGYQLTSAECYDLNSNRVADQYCHYYPENIKPKPKLQE-CNL 356
Db 172 PYSWYAPWTKCSAQAGGQVQVVECRNQLDSSAVAPHYCSGHGK----LPKQRACNT 227
QY 357 DPCPASDGYKQIMPYDLHPLRWEATPWTACSSCGGGIQSRAVSCVEEDIQGHVTSVE 416
Db 228 EPCP-----PDWVGNWSCRSCDAGVRSRVVC-----QRRVSAE 265
QY 417 EW---KCMYTPKMPIAOPCNIFDC-PKWLAQWSPCTVTTCGGLRYRVVLC--IDHKG- 468
Db 266 EKALDDSSACQPRPVPVLEACQGPCPEWATLWDSECTPCGGLRHRVLCSDAQDST 325
QY 469 MHTGCSPTKTPHIKEECIVTPCYKPKELPVEAKLPWFKQAELEGAAYSEPSFIP 528
Db 326 LPPGCHLPAKPPSPMRCNL-----RRCP-----PARWVT 355
QY 529 EAWSACTVTCGVGTQVRIVRCQVLLSFQSVADLPIDEC-EGKPKPASQACVAGPCSGEI 587
Db 356 SEWGECSQCGLGQQQRTVRC-----TSHTGQPSRECTEALRPSMQQCEA-KCDSVV 407
QY 588 PEFN-PDE 594
Db 408 PPGDGPEE 415

RESULT 11
AT56_HUMAN STANDARD; PRT; 860 AA.
ID AT56_HUMAN
AC Q9UKP5;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 01-MAR-2002 (Rel. 41, Last annotation update)
DE ADAMTS-6 precursor (bc 3.4.24.-) (A disintegrin and metalloproteinase
with thrombospondin motifs 6) (ADAM-TS 6) (ADAM-TS6).
GN ADAMTS6.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
[1]
RN SEQUENCE FROM N.A.
RP MEDLINE-99395124; PubMed-10464288;
RX

RA Hurskainen T.L., Hirohata S., Seldin M.F., Apte S.S.;
 RT "ADAM-TS5, ADAM-TS6, and ADAM-TS7, novel members of a new family of
 RL Zinc Metalloproteases.";
 CC J. Biol. Chem. 274:25555-25563(1999).
 CC -!- COFACTOR: BINDS 1 ZINC ION (BY SIMILARITY).
 CC -!- SUBCELLULAR LOCATION: SECRETED. ASSOCIATED WITH THE EXTRACELLULAR
 CC MATRIX (BY SIMILARITY).
 CC -!- TISSUE SPECIFICITY: EXPRESSED AT LOW LEVELS IN PLACENTA AND BARELY
 CC DETECTABLE IN A NUMBER OF OTHER TISSUES.
 CC -!- DOMAIN: THE SPACER DOMAIN AND THE TSP TYPE-1 DOMAINS ARE IMPORTANT
 CC FOR A TIGHT INTERACTION WITH THE EXTRACELLULAR MATRIX.
 CC -!- PTM: THE PRECURSOR IS CLEAVED BY A URIN ENDOPEPTIDASE (BY
 CC SIMILARITY).
 CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M12B.
 CC -!- SIMILARITY: CONTAINS 1 DISINTEGRIN-LIKE DOMAIN.
 CC -!- SIMILARITY: CONTAINS 2 TSP TYPE-1 DOMAINS.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; AF140674; AAD56357.1; -
 CC HSSP; P34179; 11AG.
 CC MIM; 605008;
 CC InterPro; IPR001762; Disintegrin.
 CC InterPro; IPR002870; Pep_M12B_propep.
 CC InterPro; IPR001590; Reprolysin.
 CC InterPro; IPR000884; TSP1.
 CC InterPro; IPR000130; Zn_M12Bpropep.
 CC Pfam; PF01562; Pep_M12B_propep; 1.
 CC Pfam; PF01421; Reprolysin; 1.
 CC Pfam; PF00900; tsp_1; 1.
 CC SMART; SM00209; TSP1; 1.
 CC PROSITE; PS00142; ZINC_PROPEASE; 1.
 CC PROSITE; PS00092; TSP1; 1.
 CC PROSITE; PS00427; DISINTEGRIN_1; FALSE_NEG.
 CC Repeat; Extracellular matrix.
 KW Hydrolyase; Metalloprotease; Zinc; Signal; Glycoprotein; Zymogen;
 FT SIGNAL 1 21 POTENTIAL.
 FT PROPEP 22 244 BY SIMILARITY.
 FT CHAIN 245 860 ADAMTS-6.
 FT METAL 403 404 ZINC (CATALYTIC) (BY SIMILARITY).
 FT ACT_SITE 404 404 ZINC (CATALYTIC) (BY SIMILARITY).
 FT METAL 407 407 ZINC (CATALYTIC) (BY SIMILARITY).
 FT METAL 413 413 ZINC (CATALYTIC) (BY SIMILARITY).
 FT DOMAIN 453 509 DISINTEGRIN-LIKE.
 FT DOMAIN 510 566 TSP TYPE-1 1.
 FT DOMAIN 567 668 CYS-RICH.
 FT DOMAIN 669 795 SPACER.
 FT DOMAIN 796 852 TSP TYPE-1 2.
 FT DOMAIN 68 71 POLY-ARG.
 FT DOMAIN 662 665 POLY-GLY.
 FT CARBOHYD 99 99 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 172 172 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 222 222 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 234 234 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 676 676 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 843 843 N-LINKED (GLCNAC. . .) (POTENTIAL).
 SQ SEQUENCE 860 AA; 97098 MW; E57213015D5CB2C5 CRC64;
 Query Match 5.3%; Score 506; DB 1; Length 860;
 Best Local Similarity 36.1%; Pred. No. 1.2e-23;
 Matches 117; Conservative 40; Mismatches 105; Indels 62; Gaps 9;
 QY 33 DGLWDAGPWSGSRVTCGGGASVSLRCLSL-----SKSCGGRNRYRTCSNVDCPPEAG 86
 DB 510 DGGWGPWSLWGECSRTCGGGVSSSLRCHDSPAPSGGSKYCLGERKRYRSCTDPCPLGSR 569

QY 87 DFRQAQCSAHNDYKHGQFYEMLPVSNPDNPNCSLKQAKGTLVLVLAEPKLDGTRCYT 146
 DB 570 DFRKQCADNDNPFRRGKYTNWKPITGGGVKPCALNCLAEYNYFYTERAPAVIDGTQCN 629
 QY 147 ESDMCISGLQIVGDHQLGTSVKEDNCVCGNGDSTCLVRGQYKSQLSATKSDDTTV 206
 DB 630 DSLDLCINGECKHVGCNDNLGSDAREDCRCVCGGSGTCDIEGFFNDSLPRGYME-VV 688
 QY 207 AIPGSRHRLVKGDPDHLVLETKTQGTGKNSLSGTG-----TFLVDN 251
 DB 689 QIPRGSVHI-----EVREAVMSKNYIALKSEGDYVINGAWTIDWPKFDVAG 736
 QY 252 SSVDFOKFPDK-ELLRMAGPLTADFI-----VKRNSGSASDTVOFIF 293
 DB 737 TAFHYKRPDEPSLALGPTSENILVMVLQONIGIRYKFNVPITRTGSGDNEVGFTW 796
 QY 294 -YOPIIHRWRETDFFPCSATCGG 316
 DB 797 NHQP-----WSE-----CSATCAGG 811
 RESULT 12
 ID AT58_HUMAN STANDARD; PRT; 890 AA.
 AC Q9UPF79; Q9NZS0;
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 01-MAR-2002 (Rel. 41, Last annotation update)
 DE ADAMTS-8 precursor (BC 3.4.24.-) (A disintegrin and metalloproteinase
 DE with thrombospondin motifs 8) (ADAM-TS 8) (METH-2)
 DE (METH-8).
 GN ADAMTS8 OR METH2
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Lung;
 RX MEDLINE=99367466; PubMed=10438512;
 RA Vazquez F., Hastings G., Ortega M.-A., Lane T.F., Oikemus S.,
 RA Lombardo M., Iruela-Arispe M.L.;
 RT "METH-1, a human ortholog of ADAMTS-1, and METH-2 are members of a new
 RT family of proteins with angio-inhibitory activity.";
 RT J. Biol. Chem. 274:23349-23357(1999).
 RN [2]
 SEQUENCE OF 195-440 FROM N.A.
 RX MEDLINE=20079168; PubMed=10610729;
 RA Georgiadis K.E., Hirohata S., Seldin M.F., Apte S.S.;
 RA "ADAM-TS8, a novel metalloprotease of the ADAM-TS family located on
 RT mouse chromosome 9 and human chromosome 11.";
 RL Genomics 62:312-315(1999).
 CC -!- FUNCTION: HAS ANTI-ANGIOGENIC PROPERTIES.
 CC -!- COFACTOR: BINDS 1 ZINC ION (BY SIMILARITY).
 CC -!- SUBCELLULAR LOCATION: SECRETED. ASSOCIATED WITH THE EXTRACELLULAR
 CC MATRIX (BY SIMILARITY).
 CC -!- TISSUE SPECIFICITY: HIGHLY EXPRESSED IN ADULT AND FETAL LUNG, LOWER
 CC EXPRESSION IN BRAIN, PLACENTA, HEART, STOMACH AND FETAL BRAIN AND
 CC KIDNEY.
 CC -!- DOMAIN: THE SPACER DOMAIN AND THE TSP TYPE-1 DOMAINS ARE IMPORTANT
 CC FOR A TIGHT INTERACTION WITH THE EXTRACELLULAR MATRIX.
 CC -!- PTM: THE PRECURSOR IS CLEAVED BY A URIN ENDOPEPTIDASE (BY
 CC SIMILARITY).
 CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M12B.
 CC -!- SIMILARITY: CONTAINS 1 DISINTEGRIN-LIKE DOMAIN.
 CC -!- SIMILARITY: CONTAINS 2 TSP TYPE-1 DOMAINS.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial

entities requires a license agreement (See <http://www.isb-sib.ch/announcement/> or send an email to license@isb-sib.ch).

```
CC EMBL; AF060153; RAD48081.1; -.
CC EMBL; AF175283; AAF25806.1; -.
CC HSP; P34179; I1AG.
CC MIM; 605175; -.
DR InterPro; IPR001762; Disintegrin.
DR InterPro; IPR002870; Pep_M12B_propep.
DR InterPro; IPR001590; Reprolysin.
DR InterPro; IPR000884; TSPL.
DR InterPro; IPR000130; Zn_MTPeptdse.
DR Pfam; PF01562; Pep_M12B_propep; 1.
DR Pfam; PF01421; Reprolysin; 1.
DR Pfam; PF00090; tsp.1; 2.
DR SMART; SM00209; TSPL; 2.
DR PROSITE; PS00215; ADAM_MPRO; 1.
DR PROSITE; PS00142; ZINC_PROTEASE; 1.
DR PROSITE; PS00092; TSPL; 1.
DR PROSITE; PS00427; DISINTEGRIN_1; FALSE_NEG.
KW Hydrolase; Metalloprotease; Zinc; Signal; Glycoprotein; Zymogen;
KW Repeat; Extracellular matrix; Heparin-binding.
FT SIGNAL 1 27
FT PROPEP 28 214
FT CHAIN 215 890
FT METAL 364 364
FT ACT_SITE 365 365
FT METAL 368 368
FT METAL 374 374
FT DOMAIN 439 526
FT DOMAIN 527 583
FT DOMAIN 584 690
FT DOMAIN 691 832
FT DOMAIN 833 890
FT DOMAIN 202 205
FT CARBOHYD 345 345
FT CARBOHYD 401 401
FT CARBOHYD 466 466
FT CARBOHYD 491 491
FT CARBOHYD 600 600
FT CONFLICT 195 195
FT CONFLICT 413 440
SQ SEQUENCE 890 AA; 96671 MW; 57D70E03D5739D3 CRC64;

Query Match 5.2%; Score 495.5; DB 1; Length 890;
Best Local Similarity 32.4%; Pred. No. 5.4e-23;
Matches 124; Conservative 49; Mismatches 135; Indels 75; Gaps 15;

QY 33 DGLWDAGPWSRCSRTCGGASYSLRCL-----SSKSEGRNIRYRFTCSNVDCPPPEAG 86
DB 527 DSGWAPGPGWCSRTCGGVQFSHRECKDPEPQNGRYCLGRRARYQSCHTECEPPDGK 586
QY 87 DPAQOC---SAHNDVKKHGYEWLP--VSNPDNPNCSLKCAKGTLLVVELAPKVLVG 141
DB 587 SFEQOCERYNAYNTDMDGNLLQWPKYAGVSPDRCKLFCRARGSEFKVFPEAKVIDG 646
QY 142 TRCYTESLDMCISGLCQIQVCDHOLGSTVKEDNCVCGNDGSTCRLVRGOYKSOLSATKS 201
DB 647 TLCGPETLAICVRGQCVKAGCDHVDVSPRKLDRKCGCGGKSGKRVSG-----SLTPT 700
QY 202 D---DFVVAIPYGSRII-----RLVLKGPDLHLYETKTLQGTGKGENSLSS 243
DB 701 NYGYNDIVIPAGATNIDVQKQSHPGVQNDGNYLAKTADGQYLLNGNLAIISAIEQDILV 760
QY 244 TGTFLVDNDSV-----DFQKPPDK---EILRMAG-----PLRADFIVKIR 280
DB 761 KGTILKSGSIATLERLQSFRLPELTVQLLTVPGEVFPKVKYTFVFPNDVDFSMQ-- 818
QY 281 NSSGASDTVQFIYQPLIH--RWRETFPCSAFCGGGYOLTSAECYDLNRNRVVADQYCH 339
DB 819 -SSKERATTNII--QPLLHAQWVLGWSGSSICGAGQWRVTVECRDPSCQ---ASATCN 872
```

```
QY 340 --YYPENIKPKPKLOECNLDPCP 360
DB 873 KALPEDAKP-----CESQLCP 889

RESULT 13
AT89_MOUSE
ID AT89_MOUSE STANDARD; PRT; 905 AA.
AC P57110;
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 01-MAR-2002 (Rel. 41, Last annotation update)
DE ADAMTS-8 precursor (EC 3.4.24.-) (A disintegrin and metalloproteinase
DE with thrombospondin motifs 8) (ADAM-TS 8) (ADAM-TS8) (METH-2).
GN ADAMTS8.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20079168; PubMed=10610729;
RA Georgiadis K.E., Hirohata S., Seldin M.F., Apte S.S.;
RT "ADAM-TS8, a novel metalloprotease of the ADAM-TS family located on
RL mouse chromosome 9 and human chromosome 11.";
CC Genomics 62:312-315(1999).
CC -!- FUNCTION: HAS ANTI-ANGIOGENIC PROPERTIES (BY SIMILARITY).
CC -!- COFACTOR: BINDS 1 ZINC ION (BY SIMILARITY).
CC -!- SUBCELLULAR LOCATION: SECRETED, ASSOCIATED WITH THE EXTRACELLULAR
CC MATRIX (BY SIMILARITY).
CC -!- TISSUE SPECIFICITY: EXPRESSED SPECIFICALLY IN ADULT LUNG AND HEART
CC AND LOW EXPRESSION DURING MOUSE DEVELOPMENT.
CC -!- DOMAIN: THE SPACER DOMAIN AND THE TSP TYPE-1 DOMAINS ARE IMPORTANT
CC FOR A TIGHT INTERACTION WITH THE EXTRACELLULAR MATRIX.
CC -!- PTM: THE PRECURSOR IS CLEAVED BY A FURIN ENDOPEPTIDASE (BY
CC SIMILARITY).
CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M12B.
CC -!- SIMILARITY: CONTAINS 1 DISINTEGRIN-LIKE DOMAIN.
CC -!- SIMILARITY: CONTAINS 2 TSP TYPE-1 DOMAINS.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announcement/
CC or send an email to license@isb-sib.ch).
```

```
FT DOMAIN 599 705 CYS-RICH.
FT DOMAIN 706 847 SPACER.
FT DOMAIN 848 905 TSP TYPE-1 2.
FT CARBOHYD 415 415 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 415 415 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 480 480 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 506 506 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 615 615 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 905 AA; 98879 MW; 124D4132B33A0ACE CRC64;

Query Match 4.9%; Score 474.5; DB 1; Length 905;
Best Local Similarity 30.3%; Pred. No. 1.1e-21;
Matches 117; Conservative 53; Mismatches 142; Indels 67; Gaps 16;

QY 33 DGLDANGPWSCESTRCTGGGASYSLRRC-----LSSKSCGGRNIRYRTCSNVDCPPEAG 86
DB 542 DGDWGPWRPWCSCRTGCGGQIOFSNRECDNPMFQNGRFLGERVKYQSCNTECPNGK 601
QY 87 DFRACQCSAHNDVKH---HGOFEWMLPVSN--DPDNPCLSKCAQAGTTLVVELAPKVLGD 141
DB 602 SFREOCEKYNAYNTDLDGNFLQWPVKYSGVSPDRCKLFCRARGSEFKVFEAKVIDG 661
QY 142 TRCYTESLDMCISGLCQIVGCDHQLGSGVCGNGDSTCRILVRGOYKSQLSATKS 201
DB 662 TLGCGPTLISVRCQCVKAGCDHVNPKKLDKCGVCGKGTACRKISGSP-TPEFYGVN 720
QY 202 DDTVAIPYGRSHIRLVKLG-----PDHLYLETKLQG---TKGENSLSS-----TGT 246
DB 721 D--IVTIPAGATNIDVKORSGHPGVNDGSYLAKTANGQYLLNGNLATSAIEQDILVKGT 778
QY 247 FLVNDSSV-----DFOKFPDK---EILRWAG-----PLTADFIVKIRNSG 283
DB 779 ILKYSGMTALERLOSFALEPLTVQLTVSGVEFPKRYTFVFPNDMDFSVQNSKER 838
QY 284 SADSTVQIFQPIHWRRETFDFPCSATCGGYQLTSAECYDLRSNRVADQYCH--YY 341
DB 839 AVTNIIQSL--PSA-EWVLGDWSECPSTCRGSMORRIVECDPSGQ---ASDTCDEALK 891
QY 342 PENIKPKKLOECNLDPCP 360
DB 892 PEDAKP-----CGSQPCP 904

RESULT 14
AT55_MOUSE
ID AT55_MOUSE STANDARD; PRT; 930 AA.
AC Q3R001;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 01-MAR-2002 (Rel. 41, Last annotation update)
DE ADAMTS-5 precursor (EC 3.4.24.-) (A disintegrin and metalloproteinase
DE with thrombospondin motifs 5) (ADAM-TS 5) (ADAM-TS5) (Aggrecanase-2)
DE (ADMP-2) (Implantin).
DE ADAMTS5.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=99395124; PubMed=10464288;
RA Hurskainen T.L., Hirohata S., Seidlin M.F., Apte S.S.;
RT "ADAM-TS5, ADAM-TS6, and ADAM-TS7, novel members of a new family of
RT zinc metalloproteinases."
RL J. Biol. Chem. 274:25555-25563(1999).
CC -!- FUNCTION: CLEAVES AGGECAN, A CARTILAGE PROTEOGLYCAN, AND MAY BE
CC INVOLVED IN ITS TURNOVER, MAY PLAY AN IMPORTANT ROLE IN THE
CC DESTRUCTION OF AGGECAN IN ARTHRITIC DISEASES. MAY PLAY A ROLE IN
CC PROTEOLYTIC PROCESSING MOSTLY DURING THE PERI-IMPLANTATION PERIOD.
CC -!- CATALYTIC ACTIVITY: CLEAVES AGGECAN AT THE 392-GLU-|ALA-393
CC SITE.
CC -!- COFACTOR: BINDS 1 ZINC ION (BY SIMILARITY).
CC -!- SUBCELLULAR LOCATION: SECRETED. ASSOCIATED WITH THE EXTRACELLULAR
```

```
CC MATRIX (BY SIMILARITY).
CC -!- DEVELOPMENTAL STAGE: EXPRESSED SPECIFICALLY IN THE PERI-
CC IMPLANTATION PERIOD IN EMBRYO AND TROPHOBLAST AND AT LOW OR
CC UNDETECTABLE LEVEL THEREAFTER.
CC -!- DOMAIN: THE SPACER DOMAIN AND THE TSP TYPE-1 DOMAINS ARE IMPORTANT
CC FOR A TIGHT INTERACTION WITH THE EXTRACELLULAR MATRIX.
CC -!- PTM: THE PRECURSOR IS CLEAVED BY A FURIN ENDOPEPTIDASE (BY
CC SIMILARITY).
CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M12B.
CC -!- SIMILARITY: CONTAINS 1 DISINTEGRIN-LIKE DOMAIN.
CC -!- SIMILARITY: CONTAINS 2 TSP TYPE-1 DOMAINS.
CC -----
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AF140673; AAD56356.1; -.
CC MGD; MGI:1346321; Adamts5.
CC InterPro; IPR001762; Disintegrin.
CC InterPro; IPR002870; Pep_M12B_propep.
CC InterPro; IPR001590; Reprolysin.
CC InterPro; IPR000884; TSPl.
CC InterPro; IPR000130; Zn_M12Bpeptase.
CC Pfam; PF01562; Pep_M12B_propep; 1.
CC Pfam; PF01421; Reprolysin; 1.
CC Pfam; PF00090; tsp_1; 2.
CC SMART; SM00209; TSPl; 2.
CC PROSITE; PS00215; ADAM_MEPRO; 1.
CC PROSITE; PS50092; TSPl; 1.
CC PROSITE; PS00142; ZINC_PROTEASE; 1.
CC PROSITE; PS00427; DISINTEGRIN_1; FALSE NEG.
CC Hydrolase; Metalloprotease; Zinc; Signal; Glycoprotein; Zymogen;
CC Repeat; Extracellular matrix.
CC SIGNAL 1 21 POTENTIAL.
CC PROPEP 22 261 POTENTIAL.
CC CHAIN 262 930 ADAMTS-5.
CC SITE 209 209 CYSTEINE SWITCH (POTENTIAL).
CC METAL 410 410 ZINC (CATALYTIC) (BY SIMILARITY).
CC ACT_SITE 411 411 BY SIMILARITY.
CC METAL 414 414 ZINC (CATALYTIC) (BY SIMILARITY).
CC METAL 420 420 ZINC (CATALYTIC) (BY SIMILARITY).
CC DOMAIN 485 566 DISINTEGRIN-LIKE.
CC DOMAIN 567 623 TSP TYPE-1 1.
CC DOMAIN 624 731 CYS-RICH.
CC DOMAIN 732 874 SPACER.
CC DOMAIN 875 930 TSP TYPE-1 2.
CC DOMAIN 41 46 POLY-ALA.
CC DOMAIN 257 261 POLY-ARG.
CC CARBOHYD 498 498 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC CARBOHYD 728 728 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC CARBOHYD 802 802 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC CARBOHYD 807 807 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC SEQUENCE 930 AA; 101780 MW; 84DE84B26170D4DC CRC64;

Query Match 4.9%; Score 469.5; DB 1; Length 930;
Best Local Similarity 30.3%; Pred. No. 2.3e-21;
Matches 119; Conservative 51; Mismatches 142; Indels 81; Gaps 15;

QY 34 GLDANGPWSCESTRCTGGGASYSLRRC-----LSSKSCGGRNIRYRTCSNVDCPPEAGD 87
DB 568 GNGSWGPGVCGSRSCGQVGFAYRHCNPNAPRNSRYCTGKRIYRSCSVTPCPNGKS 627
QY 88 FRAQCQCSAHN---DVKHGQFYEWLP--VSNPDNPCLSKCAQAGTTLVVELAPKVLGD 141
DB 628 FRFQCEAKNGYSDAKGVTFVWEPKYGVLPAADVCKLTKRAKGTGYVWVSPKVTGD 687
QY 142 TRCYTESLDMCISGLCQIVGCDHQLGSGVCGNGDSTCRILVRGOYKSQLSATKS 201
```

688 TCRPYSNSVVRGRCVTRGCDGIIGSKLYQDKGCVGSGDSSCTKIIGTFNKK----SKG 744
 QY 202 DFWVAIPYGSRHRL-----VLKPDHLYLTKTFLQCTKGNSLSSTGTFLVDNSSV 254
 Db 745 YDVRVPEGATHIKVQFAKQDTPREPAYLAKKKI-----GEYLN--GKYMISVS-- 795
 QY 255 DFOKFPDKILRMAGPLTADFIKIRNSGSADSTQVIFYQPIHWRWTFPPPCATG 314
 Db 796 -----ETIINDINGTV-----MNYSG-----WSHRDF-----LHG 820
 QY 315 GGYQLTS-----ACEYDLSRNRVADQYCHYPENIKPKPKLOECNLDPCASDGYKQIMP 370
 Db 821 MGSATREILLVQTLADPTKALGVRSFFV-----PKTTQKYN-----SVISHGSNKVGP 872
 QY 371 YDLYHPLPWEATPWTACSSCGGGIOSRAVSC 403
 Db 873 HSTQ---LOWTGPWLACSRICDTGWHTRVQC 902
 RESULT 15
 AT55 HUMAN STANDARD; PRT; 930 AA.
 AC Q9UNAO; Q9UNP2;
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 01-MAR-2002 (Rel. 41, Last annotation update)
 DE ADAMTS-5 precursor (EC 3.4.24.-) (A disintegrin and metalloproteinase
 with thrombospondin motifs 5) (ADAM-TS 5) (ADAM-TS5) (Aggrecanase-2)
 DE (ADMP-2) (ADAM-TS 11).
 DE ADAMTS OR ADMP2 OR ADAMTS11.
 GN Homo sapiens (Human).
 OS Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID:9606;
 RN [1]
 RN SEQUENCE FROM N.A.
 RC TISSUE=Liver;
 RX MEDLINE=99367476; PubMed=10438522;
 RA Abbaszade I., Liu R.-Q., Yang F., Rosenfeld S.A., Ross O.H.,
 Link J.R., Ellis D.M., Tortorella M.D., Pratta M.A., Hollis J.M.,
 Wynn R., Duke J.B., George H.J., Hillman M.C. Jr., Murphy K.,
 Wiswall B.H., Copeland R.A., Decicco C.P., Bruckner R., Nagase H.,
 Ito Y., Newton R.C., Magolda R.L., Trzaskos J.M., Hollis G.F.,
 Arner E.C., Burn T.C.;
 RA "Cloning and characterization of ADAMTS11, an aggrecanase from the
 ADAMTS family.";
 RT J. Biol. Chem. 274:23443-23450(1999).
 RN [2]
 RN SEQUENCE FROM N.A.
 RX MEDLINE=20289799; PubMed=10830953;
 RA Hattori M., Fujiyama A., Taylor T.D., Watanabe H., Yada T.,
 Park H.-S., Toyoda A., Ishii K., Totoki Y., Choi D.-K., Soeda E.,
 Ohki M., Takeda T., Sakaki Y., Taudien S., Blechschmidt K., Polley A.,
 Menzel U., Delabar J., Kumpf K., Lehmann R., Patterson D.,
 Reichwald K., Rump A., Schillhabel M., Schudy A., Zimmermann W.,
 Rosenthal A., Kudo J., Shibuya K., Kawasaki K., Asakawa S.,
 Shintani A., Sasaki T., Nagamine K., Mitsuyama S., Antonarakis S.E.,
 Minoshima S., Shimizu N., Nordstieck G., Hornischer K., Brandt P.,
 Scharfe M., Schoen O., Desario A., Reichelt J., Kauer G., Bloeker H.,
 Ramser J., Beck A., Klages S., Hennig S., Rieselmann L., Dagand E.,
 Wehrmeyer S., Borzym K., Gardiner K., Nizetic D., Francis F.,
 Lehrach H., Reinhardt R., Vaspo M.-L.;
 RA "The DNA sequence of human chromosome 21.";
 RT Nature 405:311-319(2000).
 RN [3]
 RN SEQUENCE OF 413-930 FROM N.A.
 RP TISSUE=Fetal brain;
 RC MEDLINE=99395124; PubMed=10464288;
 RA Hurskainen T.L., Hirohata S., Seldin M.F., Apte S.S.;
 RA "ADAM-TS6, and ADAM-TS7, novel members of a new family of
 zinc metalloproteases.";
 RT J. Biol. Chem. 274:25555-25563(1999).
 CC [1]- FUNCTION: CLEAVES AGGREGAN, A CARTILAGE PROTEOGLYCAN, AND MAY BE

INVOLVED IN ITS TURNOVER. MAY PLAY AN IMPORTANT ROLE IN THE
 DESTRUCTION OF AGGREGAN IN ARTHRITIC DISEASES. MAY PLAY A ROLE IN
 PROTEOLYTIC PROCESSING MOSTLY DURING THE PERI-IMPLANTATION PERIOD.
 -1- CATALYTIC ACTIVITY: CLEAVES AGGREGAN AT THE 392-GLU-|-ALA-393
 SITE.
 -1- COFACTOR: BINDS 1 ZINC ION (BY SIMILARITY).
 -1- SUBCELLULAR LOCATION: SECRETED. ASSOCIATED WITH THE EXTRACELLULAR
 MATRIX (BY SIMILARITY).
 -1- TISSUE SPECIFICITY: EXPRESSED AT LOW LEVEL, PRIMARILY IN PLACENTA
 BUT ALSO IN OTHER TISSUES, SUCH AS HEART AND BRAIN, AND ALSO
 CERVIX, UTERUS, BLADDER, ESOPHAGUS, RIB CARTILAGE,
 CHONDROBLASTOMA, FIBROUS TISSUE AND JOINT CAPSULE FROM AN
 ARTHRITIC PATIENT.
 -1- DOMAIN: THE SPACER DOMAIN AND THE TSP TYPE-1 DOMAINS ARE IMPORTANT
 FOR A TIGHT INTERACTION WITH THE EXTRACELLULAR MATRIX.
 -1- PTM: THE PRECURSOR IS CLEAVED BY A URIN ENDOPEPTIDASE (BY
 SIMILARITY).
 -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M12B.
 -1- SIMILARITY: CONTAINS 1 DISINTEGRIN-LIKE DOMAIN.
 -1- SIMILARITY: CONTAINS 2 TSP TYPE-1 DOMAINS.

 THIS SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).

 EMBL; AF142099; AAD49577.1; -;
 EMBL; AP011698; BAA95504.1; -;
 EMBL; AP011697; BAA95503.1; -;
 EMBL; AF141293; AAF02493.1; -;
 HSSP; Q9PW35; 1BUD.
 MIM; 605007; -;
 InterPro; IPR001762; Disintegrin.
 InterPro; IPR001590; Reprolysin.
 InterPro; IPR000884; TSP1.
 InterPro; IPR000130; Zn_Mipectase.
 Pfam; PF01421; Reprolysin; 1.
 Pfam; PF00090; tsp_1; 2.
 SMART; SM00209; TSP1; 2.
 PROSITE; PS0215; ADAM_MEPRO; 1.
 PROSITE; PS00142; ZINC_PROTEASE; 1.
 PROSITE; PS0092; TSP1; 1.
 PROSITE; PS00427; DISINTEGRIN_1; FALSE_NEG.
 KW Hydrolase; Metalloprotease; zinc; Signal; Glycoprotein; Zymogen;
 KW Repeat; Extracellular matrix.
 FT SIGNAL 1 16 POTENTIAL.
 FT PROPEP 17 261
 FT CHAIN 262 930 ADAMTS-5.
 FT SITE 209 209 CYSTEINE SWITCH (POTENTIAL).
 FT METAL 410 410 ZINC (CATALYTIC) (BY SIMILARITY).
 FT ACT_SITE 411 411 BY SIMILARITY.
 FT METAL 414 414 ZINC (CATALYTIC) (BY SIMILARITY).
 FT METAL 420 420 ZINC (CATALYTIC) (BY SIMILARITY).
 FT DOMAIN 485 566 DISINTEGRIN-LIKE.
 FT DOMAIN 567 623 TSP TYPE-1 1.
 FT DOMAIN 624 731 SPACER.
 FT DOMAIN 732 874 CYS-RICH.
 FT DOMAIN 875 930 TSP TYPE-1 2.
 FT DOMAIN 37 41 POLY-ALA.
 FT DOMAIN 257 261 POLY-ARG.
 FT CARBOHYD 498 498 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 728 728 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 802 802 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 807 807 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CONFLICT 138 138 A -> G (IN REF. 2).
 FT CONFLICT 614 614 R -> H (IN REF. 3).
 FT CONFLICT 692 692 P -> L (IN REF. 2).
 FT SEQUENCE 930 AA; 101715 MW; B64281502F28193B CRC64;

Search completed: July 24, 2002, 04:28:19
Job time: 352 sec

OM of: US-10-044-807-2 to: GenEmbl.* out_format : pfs
Date: Jul 24, 2002 6:35 AM
About: Results were produced by the GenCore software, version 4.5,
Copyright (c) 1993-2000 CompuGen Ltd.

Command line parameters:
-MODEL=framet_p2n.model -DEV=xlh
-O/cg2_1/JUSPTO_spool/US10044807/runat_22072002_153741_13656/app_query.fasta_1.1847
-DB=GenEmbl -QFMT=fastap -SUFFIX=tge -GAPOP=12.000 -GAPEXT=4.000
-MINMATCH=0.100 -LOOPEXT=0.000 -GAPEXT=0.000 -GAPEXT=4.500
-OGAPEXT=0.050 -GAPOP=10.000 -XGAPEXT=0.500 -GAPOP=6.000
-XGAPEXT=7.000 -XGAPOP=10.000 -YGAPEXT=0.500 -DELOP=6.000
-DELEXT=7.000 -START=1 -MATRIX=blomsu62 -TRANS=human40.cdi
-LIST=45 -DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0
-ALIGN=15 -MODE=LOCAL -OUTFMT=pfs -NORM=ext -HEAPSIZ=500
-MINLEN=0 -MAXLEN=200000000 -USER=US10044807_ECGN1_1.6457
-NCPU=6 -ICPU=3 -LONGLOG -DEV_TIMEOUT=120 -WARN_TIMEOUT=30
-NO_XLPXY -WAIT -THREADS=1

Search information block:
Query: US-10-044-807-2
Query length: 1762
Database: GenEmbl.*
Database sequences: 1797656
Database length: 187333701
Search time (sec): 4817.680000

score_list:

Sequence	Strd Orig	zScore	EScore	Len	Documentation
gb_pat:AX224809	+ 3956.50	2545.72	2.0e-133	5076	! AX224809 Sequence 1 from Patent
gb_pat:AX235356	+ 3953.50	2540.98	3.6e-133	7260	! AX235356 Sequence 3 from Patent
gb_pat:AX224823	+ 3762.50	2421.36	1.7e-126	4854	! AX224823 Sequence 15 from Patent
gb_pat:AX224825	+ 3673.00	2359.34	4.7e-123	8578	! AX224825 Sequence 17 from Patent
gb_pr:AF176313	+ 2910.00	1881.11	2.0e-96	1803	! AF176313 Homo sapiens ADAM-TS
gb_pat:AX224819	+ 2809.50	1813.81	1.1e-92	2538	! AX224819 Sequence 11 from Patent
gb_pr:AF237652	+ 2625.50	1695.57	4.4e-86	2523	! AF237652 Homo sapiens a disint
gb_pat:AX224821	+ 2615.50	1689.82	9.2e-86	2316	! AX224821 Sequence 13 from Patent
gb_pat:AX235354	+ 2597.50	1670.37	1.1e-84	6294	! AX235354 Sequence 1 from Patent
gb_pat:AX235410	+ 2597.50	1670.37	1.1e-84	6303	! AX235410 Sequence 57 from Patent
gb_pr:AF251058	+ 2558.00	1655.68	7.3e-84	1616	! AF251058 Homo sapiens clone 2
gb_pat:AX224815	+ 2425.50	1568.17	5.5e-79	2175	! AX224815 Sequence 7 from Patent
gb_pat:AX149471	+ 2346.50	1515.09	5.0e-76	2912	! AX149471 Sequence 27 from Patent
gb_pat:AX224817	+ 2231.50	1444.30	4.4e-72	1953	! AX224817 Sequence 9 from Patent
gb_pr:AL591423	+ 2041.00	1295.71	8.3e-64	54193	! AL591423 Human DNA sequence 4
gb_pr:AC011701	- 2041.00	1295.23	3.2e-63	205510	! AC011701 Homo sapiens 12 BAC
gb_pat:AX235377	+ 1629.00	1049.42	4.3e-50	5107	! AX235377 Sequence 24 from Patent
gb_pat:AX235379	+ 1629.00	1049.42	4.3e-50	5107	! AX235379 Sequence 26 from Patent
gb_pat:AX235411	+ 1629.00	1049.42	4.3e-50	5107	! AX235411 Sequence 58 from Patent
gb_pr:AB033059	+ 1629.00	1049.42	4.3e-50	5107	! AB033059 Homo sapiens mRNA for
gb_pat:AX235413	+ 1625.00	1046.85	6.9e-47	1341	! AX235413 Sequence 60 from Patent
gb_pat:AX224811	+ 1525.00	993.07	5.0e-47	1341	! AX224811 Sequence 5 from Patent
gb_pat:AX224813	+ 1331.00	869.78	4.4e-40	1119	! AX224813 Sequence 3 from Patent
gb_in:AY061825	+ 1280.00	827.80	9.5e-38	3604	! AY061825 Drosophila melanogast
gb_ro:AF314171	+ 1052.50	680.22	1.6e-29	4271	! AF314171 Mus musculus papilin
gb_pat:AX319857	+ 1004.00	646.62	1.2e-27	5808	! AX319857 Sequence 21 from Patent
gb_pat:AX319858	+ 999.00	645.38	1.4e-27	4518	! AX319858 Sequence 22 from Patent
gb_pr:AB037733	+ 975.00	628.94	1.1e-26	5139	! AB037733 Homo sapiens mRNA for
gb_in:AF205357	+ 944.00	606.47	2.0e-25	7100	! AF205357 Drosophila melanogast
gb_pr:AB072736	+ 921.50	595.39	8.4e-25	4614	! AB072736 Macaca fascicularis t
gb_pat:HS250725	+ 894.00	576.90	9.0e-24	5115	! HS250725 Homo sapiens mRNA for
gb_pat:ES5265	+ 887.00	572.49	1.6e-23	5061	! ES5265 Novel metalloprotease a
gb_in:AE003765	- 886.00	571.40	1.8e-23	5353	! AE003765 Sequence 2 from Patent
gb_in:AE003765	- 885.00	540.89	9.1e-22	239171	! AE003765 Drosophila melanog
gb_hlg:AC014412	+ 884.00	548.25	3.6e-22	86446	! AC014412 Drosophila melanog
gb_in:AC009213	+ 883.50	542.05	7.9e-22	182525	! AC009213 Drosophila melanog
gb_in:AF043121	+ 877.50	566.17	3.6e-23	5199	! AF043121 Haemochus contortus
gb_in:AF078161	+ 863.00	550.95	2.5e-22	11008	! AF078161 Manduca sexta lacun
gb_pr:AF414401	+ 859.50	555.65	1.4e-22	4550	! AF414401 Homo sapiens ADAMTS-3
gb_pr:AB069698	+ 859.50	554.98	1.5e-22	4950	! AB069698 Homo sapiens VWF-CP(A

gb_pr:AY055376 + 858.50 554.92 1.5e-22 4597 ! AY055376 Homo sapiens von W
gb_pr:HS235314 + 850.50 550.52 2.7e-22 4185 ! AJ305314 Homo sapiens mRNA
gb_pr:HS240810 + 850.50 550.35 2.7e-22 4278 ! AJ420810 Homo sapiens mRNA
gb_in:CEP13H10 - 822.50 516.88 2.0e-20 30601 ! Z69361 Caenorhabditis eleg
gb_pat:AX319851 + 817.00 530.82 3.3e-21 3312 ! AX319851 Sequence 15 from P
seq_name: gb_pat:AX224809
seq documentation_block:
LOCUS AX224809 5076 bp DNA linear PAT 10-SEP-2001
DEFINITION Sequence 1 from Patent WO0161011.
ACCESSION AX224809
VERSION AX224809.1 GI:15554909
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 5076)
AUTHORS Donoho,G., Scoville,J., Turner,C.A., Friedrich,G., Zambrowicz,B.
and Sands,A.T.
TITLE Novel human thrombospondin repeat proteins and polynucleotides
encoding the same
JOURNAL Patent: WO 0161011-A 1 23-AUG-2001;
Lexicon Genetics Incorporated (US)
FEATURES
Location/Qualifiers
source 1..5076
/organism="Homo sapiens"
/db_xref="taxon:9606"
BASE COUNT 1384 a 1227 c 1338 g 1127 t
ORIGIN

alignment_scores:
Quality: 3956.50 Length: 1775
Ratio: 3.248 Gaps: 31
Percent Similarity: 68.620 Percent Identity: 43.324
alignment_block:
US-10-044-807-2 x AX224809 ..
Align seg 1/1 to: AX224809 from: 1 to: 5076
22 SerSerArgThrAlaArgSerGluGluAspArgAspGlyLeuTrpAspAl 38
:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||
190 ACCTCAGAAACACACGCTTCAGATGAAGCAAGATGGCAACTGGGATGC 239
||||| |||||:::|||||:::|||||:::|||||:::|||||
38 aTrpGlyProTrpSerGluCysSerArgThrCysGlyGlyAlaSerT 55
||||| |||||:::|||||:::|||||:::|||||:::|||||
240 TTGGGGCGACTGGAGTGACTGCTCCGGACCTGTGGGGGAGGAGCATCAT 289
55 yrSerleuArgArgCysLeuSerSerSerSerCysGluGlyArgAsnIle 71
|||||:::|||||:::|||||:::|||||:::|||||:::|||||
290 ATTCTCTGGGAGATGTTTGACTGGAAGGAAATTTGTGAGGGCAGAACATT 339
72 ArgTyrArgThrCysSerAsnValAspCysProGluAlaGlyAspPh 88
|||||:::|||||:::|||||:::|||||:::|||||:::|||||
340 CGGTACACACATGACCAATCATGACGCCCTCCAGATGCAGAGATT 389
88 eArglaGlnGlnCysSerAlaHisAsnAspValLysHisGlyGlnP 105
|||||:::|||||:::|||||:::|||||:::|||||:::|||||
390 CAGAGCCCGAGCAGTGTCTCAGCTACAAATGATGTCAGATATCAGGGGCATT 439
105 heTyrGluTrpLeuProValSerAsnAspProAspAsnProCysSerLeu 121
:::|||||:::|||||:::|||||:::|||||:::|||||
440 ATATGAATGGCTTCCACGATATATGATCTCTGCTGCCCGCTGTGCATC 489
122 LysCysGlnAlaLysGlyThrThrLeuValValGluLeuAlaProLysVa 138
|||||:::|||||:::|||||:::|||||:::|||||:::|||||
490 AAGTGTCTATGCACAGACAAACAACTTGGTGGGAGCTGGCACCTAAGGT 539
138 lleuAspGlyThrArgCysTyrThrGluSerLeuAspMetCysIleSerG 155
|||||:::|||||:::|||||:::|||||:::|||||:::|||||

[illegible]

455 lyLeuArgTyrArgValValLeuCysIleAspHisArgGlyMetHisThr 471
1490 GGTACGGTACCGGGTGTCTGTGTATTAAACACCCGCGAGACGATCT
472 GlyGlyCysSerProLysThrLysProHisIleLysGluCysIleVal 488
1540 GGGGGCTCAATCCAACTCAAGTTACATCAAGAAAGATGTGTGCAT 1589
488 lProThrProCysTyrLysProLysGluLysLeuProValGluAlaLysL 505
1590 TCCCATCCGGTATTAAACCAAAAGAAAAGTCCAGTGGAAAGCAAAAT 1639
505 euProTrpPheLysGlnAlaGlnGluLeuGluGluAlaValSer 521
1640 TGCCTTGGCTGAACACAGCACAGAAGCTAGAAAGACCCAGAAATAGCAACA 1689
522 GluGluProSerPheIleProGluAlaTrpSerAlaCysThrValThrCy 538
1690 GAAGAACAACAGTTCATTCCAGAACCCCTGGTCACCCCTGCAGTACCAGGTG 1739
538 sGlyValGlyThrGlnValArgIleValArgCysGlnValLeuLeuSerP 555
1740 TGGGCGCAGGTGTGCAGTCCGCGAGGTGAAGTGCCTGCTCTCTCAT 1789
555 heSerGlnSerValAlaAspLeuProIleAspGluCysGluGlyProLys 571
1790 TCACCGAGACTGAGACTGAGTGTCCGAGGAGAGTGTGAAGGCCCAAG 1839
572 ProAlaSerGlnArg.....AlaCysTyrAlaGlyProCy 583
1840 CTGCCACCGAAGCGCCCTGCCTCTGGAAGCATGTGATGAGAGCCGGC 1889
583 sSerGlyGluIleProGluPheAsnProAspGluThrAspGlyLeuPheG 600
1890 CTCCCGAGAGCTAGACATCCCTCTCTCT..... 1917
600 lyGlyLeuGlnAspPheAspGluLeuTyrAspTrpGluTyrGluGlyPhe 616
1918GAGACAGTGNAGCAGCTTACGACTGGAGTAGCTGGGTTC 1959
617 ThrLysCysSerGluSerCysGlyGlyValGlnGluAlaValValse 633
1960 ACCCTTGCACAGCACATGCTTGGGAGCCATCAAGAAGCCATAGCAGT 2009
633 rCysLeuAsnLysGlnThrArgGluProAlaGluAsnLeuCysValT 650
2010 GTGCTTACATATCCAGACCCAGACAGTCAATGACAGCTGTGTGATA 2059
650 hrSerArgArgProGlnLeuLeuLysSerCysAsnLeuAspProCys 666
2060 TGTCTCCACCGTCTCCAGCCATGAGCCAGGCTGTAAACACAGAGCCCTGT 2109
667 ProAlaArgTrpGluIleGlyIlyTrpSerProCysSerLeuThrCysGl 683
2110 CCCCCAGTGGCATGTGGCTCTTGGGGGCCCTGCTCAGCTACTGTGG 2159
683 yValGlyLeuGlnThrArgAspValPheCysSerHisLeuLeuSerArg 700
2160 AGTTGAAATTCAGACCCGAGATGTGTACTGCCTGCAC..... 2196
700 luMetAsnGluThrValIleLeuAlaAspGluLeuCysArgGlnProLys 716
2197 ..CCAGGGGAGACCCCTGCCCTCTCTGAGGAG...TGCCGAGATGAAG 2241
717 ProSerThrValGlnAlaCysAsnArgPheAsnCysProProAlaTpy 733
2242 CCCATGCTTTACACATGCAATCAGTGTGACTGCTCCCTCTGCTGCCA 2291
733 rProAlaGlnTrpGlnProCysSerArgThrCysGlyGlyValGlnL 750
2292 CATTTGAAGATGCACAGCTGTTCAGGACTGTGGCGGGGAACACTAGA 2341

```

750 ysArgGluValLeuCysLysGlnArgMetAlaAspGlySerPheLeuGlu 766
   :|||:||||| :|||:||||| :|||:||||| :|||:||||| :|||:|||||
2342 ACAGAGAGTCACTCGCTGCGCAGCTGCTTAACGGATGGCAGCTTTTGAAT 2391
   ||| :|||:||||| :|||:||||| :|||:||||| :|||:||||| :|||:|||||
767 LeuProGluThrPheCysSerAlaSerLysProAlaCysGlnAlaCys 783
   ||| :|||:||||| :|||:||||| :|||:||||| :|||:||||| :|||:|||||
2392 CTCCTCAGATGAATTGTGCCAAGGACCAAGGATCGCTCAACAGTCCCTG 2441
   ||| :|||:||||| :|||:||||| :|||:||||| :|||:||||| :|||:|||||
783 slsLysAspLysCysProSerGluTrpLeuLeuSerAspTrpThrGluC 800
   ||| :|||:||||| :|||:||||| :|||:||||| :|||:||||| :|||:|||||
2442 TGCCAGGACAGACTGCTCTCCACATTTAGCTCTGGGAGACTGGTCAAGT 2491
   ||| :|||:||||| :|||:||||| :|||:||||| :|||:||||| :|||:|||||
800 ysSerThrSerCysGluGlyThrGlnThrArgSerAlaLeuCysArg 816
   ||| :|||:||||| :|||:||||| :|||:||||| :|||:||||| :|||:|||||
2492 GTTCTGTCAGTTGTGGTGTGAATCCAGAGAAAGCAAGCGTGTCTCAA 2541
   ||| :|||:||||| :|||:||||| :|||:||||| :|||:||||| :|||:|||||
817 LysMetLeuLysThrGlyLeuSerThrValValAsnSerThrLeuCysPr 833
   ||| :|||:||||| :|||:||||| :|||:||||| :|||:||||| :|||:|||||
2542 AGGCTGGCAGCCAAAGGTGGCGCATCCCTCAGTGAGATGATGTGCAG 2591
   ||| :|||:||||| :|||:||||| :|||:||||| :|||:||||| :|||:|||||
833 oProLeuProPheSerSerIleArgProCysMetLeuAlaThrCysA 850
   ||| :|||:||||| :|||:||||| :|||:||||| :|||:||||| :|||:|||||
2592 GGATCTACAGCGGTTCCTCTTGTGAAGATCTTGCAGATGCTGAGTGCA 2641
   ||| :|||:||||| :|||:||||| :|||:||||| :|||:||||| :|||:|||||
850 laArg...ProGlyArgProSerThrLys.....HisSerProHis 862
   ||| :|||:||||| :|||:||||| :|||:||||| :|||:||||| :|||:|||||
2642 GTAAATCAATCAGAGATGAAGACAAACTGGTGACGAGGTCCCGCAG 2691
   ||| :|||:||||| :|||:||||| :|||:||||| :|||:||||| :|||:|||||
863 lleAlaAlaArgLysValTyrIleGlnThrArgArgGlnArgLysIle 879
   ||| :|||:||||| :|||:||||| :|||:||||| :|||:||||| :|||:|||||
2692 ATCTCAGTGTCCAGAGATCTACATTCAGACAAAGGAGAGAACGGTAT 2741
   ||| :|||:||||| :|||:||||| :|||:||||| :|||:||||| :|||:|||||
879 uHisPheValValGlyGlyPheAlaTyrLeuLeuProLysThrAlaVal 896
   ||| :|||:||||| :|||:||||| :|||:||||| :|||:||||| :|||:|||||
2742 TAACTGACCATGGTAGCAGAGCTATTGTGTCGCCAACACATCCGTGA 2791
   ||| :|||:||||| :|||:||||| :|||:||||| :|||:||||| :|||:|||||
896 alLeuArgCysProAlaArgValArgLysProLeuIleThrTrpGlu 912
   ||| :|||:||||| :|||:||||| :|||:||||| :|||:||||| :|||:|||||
2792 TTATTAGTCCCGCGCGAGCATTCAGAAATCTCTGATCCAGTGGGAG 2841
   ||| :|||:||||| :|||:||||| :|||:||||| :|||:||||| :|||:|||||
913 LysAspGlyGlnHisLeuLeuSerSerThrHisValThrValAlaProPh 929
   ||| :|||:||||| :|||:||||| :|||:||||| :|||:||||| :|||:|||||
2842 AAGGATGGCGCTTGCAGAGACTCCAAACGGCTTGGCATCACCAGTC 2891
   ||| :|||:||||| :|||:||||| :|||:||||| :|||:||||| :|||:|||||
929 eClyTyLeuLysIleHisArgLeuLysProSerAspAlaGlyValTyr 946
   ||| :|||:||||| :|||:||||| :|||:||||| :|||:||||| :|||:|||||
2892 AGGCTCACTAAATCCACGGCTCTTGTGCCCGCGACATCGCGTGTACC 2941
   ||| :|||:||||| :|||:||||| :|||:||||| :|||:||||| :|||:|||||
946 hrCysSerAlaGlyProAlaArgGluHisPheValIleLysLeuIleCly 962
   ||| :|||:||||| :|||:||||| :|||:||||| :|||:||||| :|||:|||||
2942 GGTGCATTGCAGGCTCTGCACAGGAACAGTTGTGCTCAAGCTCATTTGGT 2991
   ||| :|||:||||| :|||:||||| :|||:||||| :|||:||||| :|||:|||||
963 GlyAsnArgLysLeuValAlaArgProLeuSerProArgSerGluGlu 979
   ||| :|||:||||| :|||:||||| :|||:||||| :|||:||||| :|||:|||||
2992 ACTGACAAACGGGTCTATCGCAGCGCCA..... 3018
   ||| :|||:||||| :|||:||||| :|||:||||| :|||:||||| :|||:|||||
979 uValLeuAlaGlyArgLysGlyProLysGluAlaLeuGlnThrHisL 996
   ||| :|||:||||| :|||:||||| :|||:||||| :|||:||||| :|||:|||||
3019 .....GCCCTCAGGAGCGCTATGAGGGAATATCCCTGGATGGACCACA 3061
   ||| :|||:||||| :|||:||||| :|||:||||| :|||:||||| :|||:|||||
996 ysHisGlnAsnGlyIlePheSerAsnGlySerLysAlaGluLysArgGly 1012
   ||| :|||:||||| :|||:||||| :|||:||||| :|||:||||| :|||:|||||
3062 GCGAAGCCATAGTTGGGACTCACATGGCACAAA.....ATGAGGCAA 3105
   ||| :|||:||||| :|||:||||| :|||:||||| :|||:||||| :|||:|||||
1013 LeuAlaAlaAsnProGlySerArgTyr.....AspAspLeuValSerAr 1027
   ||| :|||:||||| :|||:||||| :|||:||||| :|||:||||| :|||:|||||
3106 ATGTGGAATAACAAAATGACCTTATCTGGATGATGACCACATTATGTA 3155
   ||| :|||:||||| :|||:||||| :|||:||||| :|||:||||| :|||:|||||
1027 g.....LeuLeuGluGlnGlyTyrTrpProGlyGluL 1038
   ||| :|||:||||| :|||:||||| :|||:||||| :|||:||||| :|||:|||||
3156 CCAGCCTTCTTGAGAGCTCTGTAGGCCCATTCGACGAATTCGACGAA 3205
   ||| :|||:||||| :|||:||||| :|||:||||| :|||:||||| :|||:|||||
1038 euLeuAlaSerTrpGluAlaGlnAspSerAlaGluArgAsnThrThrSer 1054
   ||| :|||:||||| :|||:||||| :|||:||||| :|||:||||| :|||:|||||

```

```

3206 GCACCAACTCCTGGAGTTCAAGAAATAAGCAGTTTGAAGCAGCAGTAAA 3255
   |||:||||| :|||:||||| :|||:||||| :|||:||||| :|||:|||||
1055 GluGluAspProGlyAlaGluGlnValLeuLeuHisLeuProPheThrMe 1071
   |||:||||| :|||:||||| :|||:||||| :|||:||||| :|||:|||||
3256 CAA.....GGAGCA.....TATAGCAT 3272
   |||:||||| :|||:||||| :|||:||||| :|||:||||| :|||:|||||
1071 tValThrGluGlnArgArgLeuAspPheLeuGlyAsnLeuSerGlnG 1088
   |||:||||| :|||:||||| :|||:||||| :|||:||||| :|||:|||||
3273 GGATACAGCCAG.....TTGATGAGCTGATAAGAAACATCAGTCAGC 3316
   |||:||||| :|||:||||| :|||:||||| :|||:||||| :|||:|||||
1088 InProGlu.....GluLeuArgAspLeuTyrSerLysHisLeuValAla 1102
   |||:||||| :|||:||||| :|||:||||| :|||:||||| :|||:|||||
3317 TCATGGAAACCGAGAGGTGAGGATGATCTTCGCTCCAGCTGATATAT 3366
   |||:||||| :|||:||||| :|||:||||| :|||:||||| :|||:|||||
1103 GlnLeuAlaGlnGluPheArgSerHisLeuGluHisGlnAspThrLe 1119
   |||:||||| :|||:||||| :|||:||||| :|||:||||| :|||:|||||
3367 CAGCTGGTGGCGGAATTAGCCAAGGCACACACACATGCAG..... 3411
   |||:||||| :|||:||||| :|||:||||| :|||:||||| :|||:|||||
1119 uLeuLysProSerGluArgArgThrSerProValThrLeuSerProHisL 1136
   |||:||||| :|||:||||| :|||:||||| :|||:||||| :|||:|||||
3412 .TGGCGGGCATCCAGGAAGAGACACCTCCT..... 3441
   |||:||||| :|||:||||| :|||:||||| :|||:||||| :|||:|||||
1136 ysHisValSerGlyPheSerSerSerLeuArgThrSerSerThrGlyAsp 1152
   |||:||||| :|||:||||| :|||:||||| :|||:||||| :|||:|||||
3442 .....GCTGCTCAGCTCAGA.....GGGGA 3462
   |||:||||| :|||:||||| :|||:||||| :|||:||||| :|||:|||||
1153 AlaGlyGlySerArgArgProHisArgLys..... 1163
   |||:||||| :|||:||||| :|||:||||| :|||:||||| :|||:|||||
3463 ACAGGAGTGTGTCCCAAGCTCGCATGCAAAAACTCAGGCAAGCTGAC 3512
   |||:||||| :|||:||||| :|||:||||| :|||:||||| :|||:|||||
1164 .....ProThrIleLeuArgLysIleSerAlaAlaGlnG 1175
   |||:||||| :|||:||||| :|||:||||| :|||:||||| :|||:|||||
3513 ATTCAGCGGAAGACCTGTCTCATGAGGCAA...AGCCAAACCTCCCT 3559
   |||:||||| :|||:||||| :|||:||||| :|||:||||| :|||:|||||
1175 InLeuSerAlaSerGluValVal...ThrHisLeuGlyGlnThrValAla 1190
   |||:||||| :|||:||||| :|||:||||| :|||:||||| :|||:|||||
3560 CAATTTCATTTAATAAACAATAAATCCAGGTTGGAAATACAGTATAC 3609
   |||:||||| :|||:||||| :|||:||||| :|||:||||| :|||:|||||
1191 LeuAlaSerGlyThrLeuSerValLeuLeuHisCysGluAlaIleGlyH 1207
   |||:||||| :|||:||||| :|||:||||| :|||:||||| :|||:|||||
3610 ATTCAAAAAGGACACAGGTCTCAATATATCTGTGTGACCTTATTACCC 3659
   |||:||||| :|||:||||| :|||:||||| :|||:||||| :|||:|||||
1207 sProArgProThrIleSerTrpAlaArgAsnGlyGluGlnValGlnPheS 1224
   |||:||||| :|||:||||| :|||:||||| :|||:||||| :|||:|||||
3660 CAGTGAGCCACATATACATGACCAAGGATGGAACCTTGTGTACAGCCCT 3709
   |||:||||| :|||:||||| :|||:||||| :|||:||||| :|||:|||||
1224 eAspArgIleLeuLeuGlnProAspSerLeuGlnIleLeuAlaPro 1240
   |||:||||| :|||:||||| :|||:||||| :|||:||||| :|||:|||||
3710 CAGTAAAAATAATTTGGATGGAAGTGGGAAGATACAGATCAGATCCT 3759
   |||:||||| :|||:||||| :|||:||||| :|||:||||| :|||:|||||
1241 ValGluAlaAspValGlyPheTyThrCysAsnAlaThrAsnAlaLeuGl 1257
   |||:||||| :|||:||||| :|||:||||| :|||:||||| :|||:|||||
3760 ACAAGGAAAGAACCAAGCATATATGAATGTCTGTAGCTAATCATCTTGG 3809
   |||:||||| :|||:||||| :|||:||||| :|||:||||| :|||:|||||
1257 yTyAspSerValSerIleAlaValThrLeuAlaGlyLysProLeuValL 1274
   |||:||||| :|||:||||| :|||:||||| :|||:||||| :|||:|||||
3810 TTCAGATGTGAAAGTCTTCTGTGTGTATGTCAGAGGACCTCTCATCT 3859
   |||:||||| :|||:||||| :|||:||||| :|||:||||| :|||:|||||
1274 ysThrSerArgMetThrValIleAsnThrGluLysProAlaValThrVal 1290
   |||:||||| :|||:||||| :|||:||||| :|||:||||| :|||:|||||
3860 TGTCTGTTGAAAGAAATATACCAACACAGAGACACACCATCTCTGTGT 3909
   |||:||||| :|||:||||| :|||:||||| :|||:||||| :|||:|||||
1291 AspIleGlySerThrIleLysThrValGlnGlyValAsnValThrIleAs 1307
   |||:||||| :|||:||||| :|||:||||| :|||:||||| :|||:|||||
3910 GTGCTTGAGGCGCATCGTGAGGCGACCTTGGAGCAACCGTGACCAATCC 3959
   |||:||||| :|||:||||| :|||:||||| :|||:||||| :|||:|||||
1307 nCysGlnValAlaGlyValProGluAlaGluValThrTrpPheArgAsnL 1324
   |||:||||| :|||:||||| :|||:||||| :|||:||||| :|||:|||||
3960 ATGCTCTGTAAGAGGTGCTCCCTCAGCCTAATAATACTTGGTTGAAGAG 4009
   |||:||||| :|||:||||| :|||:||||| :|||:||||| :|||:|||||
1324 ysSerLysLeu...GlySerProHisHisLeuHisGluGlySerLeuLeu 1339
   |||:||||| :|||:||||| :|||:||||| :|||:||||| :|||:|||||

```


505 euProTrrpPheLysGlnAlaGlnGluLeuGluGlyAlaAlaValSer 521
 1775 TGCCTTGCCTGAACAAGACAAGAAGACTAGAAGACACACCAAGATAGCAACA 1824

522 GluGluProSerPheLleProGluAlaTtpSerAlaCysThrValThrcy 538
 1825 GAAGAACCAACGTTTCATTCAGAACCCCTGGTCAGCTCGCATGATACACG 1874

538 sGlyValGlyThrGlnValArgIleValArgCysGlnValLeuLeuSerP 555
 1875 TGGCCGGGTGTGCAGTCCGTGAGGTGAAGTCCCGCTGTGCTCTCACAT 1924

555 heSerGlnSerValAlaAspLeuProIleAspGluCysGluGlyProLys 571
 1925 TCAGCGAGCTGAGCTGAGCTCCCGAGGAAGAGTGTGAAGGCCCAAG 1974

572 ProAlaSerGlnArg.....AlaCysTyrAlaGlyProcy 583
 1975 CTGCCACCGNACGGCCCTGCCCTCTGGAGCATGTGATGAGAGCCCGC 2024

583 sSerGlyGluLeuProGluPheAsnProAspGluThrAspGlyLeuPheG 600
 2025 CTCGCGAGAGCTAGACATCCCTCTCCCT..... 2052

600 LyGlyLeuGlnAspPheAspGluLeuTyrAspTrpGluTyrGluGlyPhe 616
 2053GAGGACAGTGGAGCATTTACGACGTGGGAGTACGCTGGGTTC 2094

617 ThrLysCysSerGluSerCysGlyGlyValGlnGluAlaValAlaValSe 633
 2095 ACCCCTTGCACAGCAACATGCTGGGAGGCCATCAAGAGCCATATGACGT 2144

633 rCysLeuAsnLysGlnThrArgGluProAlaGluGluAsnLeuCysValT 650
 2145 GTGCTTACATATCCAGACCCAGCAGACAGTCAATGACAGCTTGTGTGATA 2194

650 hrSerArgArgProGlnLeuLeuLysSerCysAsnLeuAspProCys 666
 2195 TGGTCCACCTTCCTCCACCATGATGACGAGCCGCTGTAACACAGAGCCCTGT 2244

667 ProAlaArgTrpGluLeuGlyLysTrpSerProCysSerLeuThrCysGl 683
 2245 CCCCCAGGTGGCATGTGGGCTCTTGGGGGCCCTGTCTCAGTACCTGTGTG 2294

683 vValGlyLeuGlnThrArgAspValPheCysSerHisLeuLeuSerArg 700
 2295 AGTTGGAAITTCAGACCCGAGTGTACTGGCTGCAC..... 2331

700 LuMetAsnGluThrValIleLeuAlaAspGluLeuCysArgGlnProLys 716
 2332 ..CCAGGGGAGACCCCTGCCCTCTCTCAGGAG...TGCCGAGATGAAAG 2376

717 ProSerThrValGlnAlaCysAsnArgPheAsnCysProProAlaTrpTy 733
 2377 CCCCATGTTTACAGATGCAATCAGTTGTACTGGCCCTCTGGCTGGCA 2426

733 rProAlaGlnTrpGlnProCysSerArgTrpCysGlyGlyGlyValGlnL 750
 2427 CATTTGAGNATGGCAGAGTGTTCAGGACTTGTGGCGGGGAACCTCAGA 2476

750 yArgGluValLeuCysLysGlnArgMetAlaAspGlySerPheLeuGlu 766
 2477 ACAGAAGAGTCACTGTCCGCGAGCTGTCTAAGCGATGCGAGCTTTTGAAT 2526

767 LeuProGluThrPheCysSerAlaSerLysProAlaCysGlnGlnAlaCy 783
 2527 CTCTCAGATGAAITGTGCCAGAGGCCCCNAGGCATGCTCTCACAGTCTCTG 2576

783 sLysLysAspAspCysProSerGluTrpLeuLeuSerAspTrpThrGluC 800
 2577 TGCCAGCAGACAGTCTCTCCACATTTACGTCTGGGAGACTGGTCTGAAGT 2626

800 ysSerThrSerCysGlyGluGlyThrGlnThrArgSerAlaIleCysArg 816
|||||:::|||||:::||||| ||| ||| ||||::: ||||:::
2627 CTTCTGTCACTTGGTGTTGGATCCAGACAAGAAGCAGGTGTCACAA 2676

817 LysMetLeuLysThrGlyLeuSerThrValValAsnSerThrLeuCysPr 833
:::|::: |::: |::: |::: |::: |::: |::: |::: |::: |::: |
2677 AGGCTGGCAGCCAAGGTGGCGCATCCCCCTCAGTGAGATGATGTCAG 2726

833 oProLeuProPheSerSerSerrilleAargProCysMetLeuAlaThrCysA 850
||||| ::||| ::||| ||||::: ||||::: |||:
2727 GGATCTACCAGGGTTCCTCTWTGTGAAGATCTTGCCAGATGCCCTCAGTGA 2776

850 LaArg...ProGlyArgProSerThrLys.....HisSerProHis 862
:::|::: |::: |::: |::: |::: |::: |::: |::: |::: |::: |
2777 GTAAATAACAAATCACAGATGAAGCAAACACTGGTGACAGAGGTGCCAG 2826

863 IleAlaAlaAlaArgLysValTyrlleGlnThrArgArgGlnArgLysLe 879
:::|::: |::: |::: |::: |::: |::: |::: |::: |::: |::: |
2827 ATCCTCAGTGTCCAGAGAGTCTACATTCAGAACAAAGGAGAGAGCGTAT 2876

879 uHispheValValGlyGlyPheAlaTyrlleuLeuProLysThrAlaVal 896
:::|::: |::: |::: |::: |::: |::: |::: |::: |::: |::: |
2877 TAAOCTGACCAATTGGTAGCAGAGGCTATTGCTGCCCAACACATCCGTGA 2926

896 alLeuArgCysProAlaArgValArgLysProLeulleThrTrpGlu 912
:::|::: |::: |::: |::: |::: |::: |::: |::: |::: |::: |
2927 TTATTTAAGTGCOCCTGGCAGCATCCAGAAATCTGTGATCCAGTGGAG 2976

913 LysAspGlyGlnHisLeuIleSerSerThrHisValThrValAlaProPh 929
|||||:::|||:::|||:::|||:::|||:::|||:::|||:::|||:::|||:::
2977 AAGGATGGCCGTGTGCTGCAGAACTCCAAAGCGGTGGCATCACCAAGTC 3026

929 eGlyTyrlleuLysIleHisArgLeuLysProSerAspAlaGlyValTyrt 946
||| |::| |::| |::| |::| |::| |::| |::| |::| |::| |::| |
3027 AGGCTCACTAAAATCCAGGTCTTGCTGCCCGCACATCGGCGTGTACC 3076

946 hrCysSerAlaGlyProAlaArgGluHisPheValIleLysLeuilleGly 962
||| |::| |::| |::| |::| |::| |::| |::| |::| |::| |::| |
3077 GGTGCATTCAGGCTCTGCACAGGAACAGATGTGTGCTCAAGCTCATTTGGT 3126

963 GlyAsnArgLysLeuValAlaArgProLeuSerProArgSerGluGluGl 979
|||||:::|||:::|||:::|||:::|||:::|||:::|||:::|||:::|||:::
3127 ACTGACAAACCGCTCATCGACGGCCA..... 3153

979 uValLeuAlaGlyArgLysGlyGlyProLysGluAlaLeuGlnThrHisL 996
||| |::| |::| |::| |::| |::| |::| |::| |::| |::| |::| |
3154GCCCTCAGGGAGCCTATGAGGGAATATCCTGGGATGSAACCACA 3196

996 ySHisGlnAsnGlytlePheSerAsnGlySerLysAlaGluLysArgGly 1012
:::|::: |::: |::: |::: |::: |::: |::: |::: |::: |::: |
3197 GCGAAGCCAAATAGTTGGAGTCATCGGCACAAA.....ATCAGGCAA 3240

1013 LeuAlaAlaAsnProGlySerArgTyrr.....AspAspLeuValSerAr 1027
||| |::| |::| |::| |::| |::| |::| |::| |::| |::| |
3241 ATGTGGATACAAAATGACCTTTATCTGSATGATGACCACATAGTAA 3290

1027 g.....LeuLeuGluGlnclGlyltyrppProGlyGluL 1038
; ||||| ::|::: |::: |::: |::: |::: |::: |::: |::: |::: |
3291 CCAGCGCTTCTTTCGAGAGCTCTGTTAGGCACTGCGCAATTCGCGGAA 3340

1038 euLeuAlaSerTrpGluAlaAspSerAlaGluArgAsnThrThrSer 1054
|||||:::|||:::|||:::|||:::|||:::|||:::|||:::|||:::|||:::
3341 GCACCAACTCCCTGGGAGTTGAAGATAAGCAGTTTGAAGACGAGTATAA 3390

1055 GluGluAspProGlyAlaGluGlnValLeuHisLeuProPheThrMe 1071
:::|::: |::: |::: |::: |::: |::: |::: |::: |::: |::: |
3391 CAA.....GGACA.....TATAGCAT 3407

1071 tValThrGluGlnArgArgLeuAspAspFleuLeuGlyAsnLeuSerGln 1088
||| |::| |::| |::| |::| |::| |::| |::| |::| |::| |
3408 GGATACAGCCCAAG.....TTTGATGAGCTGTGATAAGAACATGAGTCAGC 3451

1088 InProGlu.....GluLeuArgAspLeuTyrrSerLysHisLeuValAla 1102

1685 TyrGlyPheGlnSerArgArgValGluCysValHisAlaArgThrAsnLys 1701
|||||
4981 ...GGTTCCAGTCTCGGAAGTCGACTGTATCCACACAGGAGTTGCCAA 5027
1701 sAlaValProGluHisLeuCysSerTrpGlyProArgProAlaAsnTrpG 1718
| ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
5028 ACCTGTGGCCAGAGACACTGTGTACAGAAAAGAAACCAATTTCTGGC 5077
1718 lnArgCysAsnIleThrProCysGlnAsnMetGluCysArgAspThrThr 1734
: : : : : ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
5078 GGCACGTGT...CTGGGCCCTCTGTGTAGAGACTGCACAGACACACT 5124
1735 ArgTyrCysGlnLysValHisGlnLeuLysLeuCysGlnLeuSerGlnPh 1751
: : : : : ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
5125 CACTACTGTATGTTGTAAACATCTTAATTTGTGTCTCTAGACGCGTA 5174
1751 eLysSerArgCysCysGlyThrCys 1759
: : : : : ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
5175 CAACCAAGGTGTCGACGTATGT 5199

seq_name: gb_pat:AX224823
seq_documentation_block:
LOCUS AX224823 4854 bp DNA linear PAT 10-SEP-2001
DEFINITION Sequence 15 from Patent WO0161011.
ACCESSION AX224823
VERSION AX224823.1 GI:15554916
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 4854)
AUTHORS Donoho,G., Scoville,J., Turner,C.A., Friedlich,G., Zambrowicz,B.
and Sands,A.T.
TITLE Novel human thrombospondin repeat proteins and polynucleotides
encoding the same
JOURNAL Patent: WO 0161011-A 15 23-AUG-2001;
Lexicon Genetics Incorporated (US)
FEATURES
source 1..4854
/organism="Homo sapiens"
/db_xref="taxon:9606"
BASE COUNT 1331 a 1175 c 1276 g 1072 t
ORIGIN

alignment_scores:
Quality: 3762.50 Length: 1733
Ratio: 3.194 Gaps: 31
Percent Similarity: 67.975 Percent Identity: 42.527
alignment_block:
US-10-044-807-2 x AX224823 ..
Align seg 1/1 to: AX224823 from: 1 to: 4854
64 LysSerCysGluGlyArgAsnIleArgTyrArgThrCysSerAsnValas 80
: : : : : ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
94 AGGAATGTGAAGGCGACACATTCGGTACAGACATGCAGCAATCATGA 143
80 pCysProGluAlaGlyAspPheArgAlaGlnGlnCysSerAlaHisA 97
|||||
144 CTGCCCTCCAGATGCAGAAGATTTTCAGAGCCCGCAGCAGTGTCTCAGCCTACA 193
97 snAspValHisHisGlyGlnPheTyrGluTrpLeuProValSerAsn 113
|||||
194 ATGATGTCCAGTATCAGGGCGATTACTATGATGGCTTCCAGCATATAT 243
114 AspProAspAsnProCysSerLeuLysCysGlnAlaLysGlyThrThrie 130
|||||
244 GATCCTGTGCCCGCTGTGTCATCAAGTGTATGCACACAGGACAAACTT 293

130 uValValGluLeuAlaProLysValLeuAspGlyThrArgCysTyrThrG 147
|||||
294 GGTGTGGAGCTGGCACTAAGGTACTGGATGGAACTCGTTGGCAACACGG 343
147 luSerLeuAspMetCysIleSerGlyLeuCysGlnIleValGlyCysAsp 163
| ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
344 ACCTCTGGACATGTGTATCAGTGGCATCTGTACGAGCAGTGGGTGGCAT 393
164 HisGlnLeuGlySerThrValLysGluAspAsnCysGlyValCysAsnG 180
: : : : : ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
394 CGSCACTGGGAAGCAATGCCAGGAGACAACTGTGGAGTCTGTGCGG 443
180 yAspGlySerThrCysArgLeuValArgGlyGlnTyrLysSerGlnLeuS 197
|||||
444 CGATGGCTCCACCTGCGAGGCTGTACGGGCAATCAAAAGTCACACGTT 493
197 erAlaThrLysSerAspAspThrValValAlaIleProTyrGlySerArg 213
| ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
494 CTCTGGAAGAAAGAGAGAAATGTAATGCTGTCTTGGGAAGTCGA 543
214 HisIleArgLeuValLeuLysGlyProAspHisLeuTyrLeuGluThrLy 230
: : : : : ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
544 AGTGTGAGAAATTACAGTGAAGGACCTGCCACCTCTTTATTGAATCAA 593
230 sThrLeuGlnGlyThrLysGlyGluAsnSerLeuSerSerThrGlyThrP 247
|||||
594 AACACTTCAAGGAAGCAAGGAGAACACAGCTTTAACAGAGCCCGCGTCT 643
247 heLeuValAspAsnSerSerValAspPheGlnLysPheProAspLysGlu 263
| ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
644 TTGCTGTAGAANAACACACAGCTGGAATTCAGAGGGCTCCGAGAGGCAA 693
264 IleLeuArgMetAlaGlyProLeuThrAlaAspPheIleValLysIleAr 280
: : : : : ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
694 ACTTTTAAGATTCAGGACCTCTGATGCTGATTTTCATCTTCAAGACCAG 743
280 gAsnSerGlySerAlaAspSerThrValGlnPheIlePheTyrGlnProI 297
| : : : : : ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
744 GTACACTGCAGCCAAAGACAGCGTGGTTCAGTTCCTTTTACCAGCCCA 793
297 leIleHisArgTrpArgGluThrAspPhePheProCysSerAlaThrCys 313
| ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
794 TCAGTCATCAGTGGAGACAACTGACTTCTTCCCTGCACCTGTGACGTGT 843
314 GlyGlyGlyTyrGlnLeuThrSerAlaGluCysTyrAspLeuArgSerAs 330
|||||
844 GGAGGAGGTATACAGCTCAATTCGTGAATGTGTGATATCCGCTTGA 893
330 nArgValValAlaAspGlnTyrCysHisTyrTyrProGluAsnIleLysp 347
: ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
894 GAGGTAGTCTCTGACCAATATTGTCTACTACTTACCTGAAATGTAACAC 943
347 rOLysProLysLeuGlnCysAsnLeuAspProCysProLysSerAsp 363
|||||
944 CAANAACCAAACTGAAGGAATGCAGCATGGATCCCTGCCCATCAAGTGAT 993
364 GlyTyrLysGlnIleMetProTyrAspLeuTyrHisProLeuProArgTr 380
|||||
994 GGATTAAGAGATAATGCCCTATGACCACCTTCCAACTCTTCTCCGCTG 1043
380 pGluAlaThrProTrpThrAlaCysSerSerCysGlyGlyGlyLeG 397
|||||
1044 GGAACATAATCTTGGACATGCGATGTCGCTGTCCTGTCGAGGAGGATTC 1093
397 lnSerArgAlaValSerCysValGluGluAspIleGlnGlyHisValThr 413
| ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
1094 AGAGAGCGGAGCTTGTGTGTGTAGAGAAATCCATGCATGCAGAGATATTG 1143
414 SerValGluGluTrpLysCysMetTyrThrProLysMetProIleAlaG 430
: ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
1144 CAGGTGAAGAAATGGAAGTGCATGTACGCCACCAACCAAGGTATGCA 1193

wed Jul 24 11:35:42 2002

```
2910 TCTGGATGATGACCAACATTAGTAACAGACCTTTCTTGAGAGCTCTGTTAG 2959
1030 luGlnClyGlyTrpProGlyGluLeuLeuAlaSerTrpGluAlaGlnAsp 1046
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
2960 GCCACTGCGCAATTCCTGAGGAAGACCAACTCCTGGAGTGAAGAAT 3009
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1047 SerAlaGluArgAsnThrThrSerGluGluAspProGlyValaGlnVa 1063
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
3010 AAGCATTGGAACGACAGTTAAACA.....GGAGCA..... 3042
1063 lLeuLeuHisLeuProPheThrMetValThrGluGlnArgArgLeuAspA 1080
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
3043 .....TATAGCATGATACAGCCAG.....TTTGATG 3070
1080 spLeuLeuGlyAsnLeuSerGlnProGlu.....GluLeuArgAsp 1094
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
3071 AGCTGATAAGAAACATGAGTCAGCTCATGGAACCGGAGAGTCAGCGAT 3120
1095 LeuTyrSerLysHisLeuValAlaGlnLeuAlaGlnIlePheArgse 1111
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
3121 GATCTTCGTCCTCCAGCTGATATATACCTGCTGGCGGAATTAGCCAAAGC 3170
1111 rHisLeuGluHisGlnAspThrLeuLeuLysProSerGluArgThrS 1128
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
3171 ACAGCCAACACACATGCAG.....TGGCGGGGCATCCAGGAAGAGAC 3214
1128 erProValThrLeuSerProHisLysHisValSerGlyPheSerSer 1144
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
3215 CTCT.....GCTGCTCAG 3228
1145 LeuArgThrSerSerThrGlyAspAlaGlyGlyGlySerArgArgProH 1161
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
3229 CTCAGA.....GGGAAACAGGCGAGTGTGCCCAAGCTCGCA 3266
1161 sArgLys.....ProThrIleL 1167
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
3267 TGCAAAAAACTCAGCAAGCTGACATTCAGAGCGGAAAGACCTGTTCTCA 3316
1167 euArgLysIleSerAlaAlaGlnLeuSerAlaSerGluValVal... 1182
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
3317 TGAGGCAA...AGCCAACCTCCCTCAATTCATTATATAAAACAATAAAT 3363
1183 ThrHisLeuGlyGlnThrValAlaLeuAlaSerGlyThrLeuSerValle 1199
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
3364 TCCAGGATTGGAAATACAGATATACATTACAAAAGGACAGAGGTCATCAA 3413
1199 uLeuHisCysGluAlaIleGlyHisProArgProThrIleSerTrpAla 1216
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
3414 TATACTGTGTGACCTATTATACCCCGCTGAGGCCACATATATACATGGACCA 3463
1216 rgAsnGlyGluGluValGlnPheSerAspArgIleLeuLeuGlnProasp 1232
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
3464 AGGATGGAACTTGTATACAGCCCTCAGTAAATAATTTTGGATGGAAC 3513
1233 AspSerLeuGlnIleLeuAlaProValGluAlaAspValGlyPheTyrTh 1249
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
3514 GGGAGATACAGATACAGAATCCTACAAGGAAGAAACAAGCATATATGA 3563
1249 rCysAsnAlaThrAsnAlaLeuGlyTyrAspSerValSerIleAlaValT 1266
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
3564 ATGTCTGTAGCTAATCACTGTGGTTCAGATGTGGAAGTCTTCCTGTGC 3613
1266 hrLeuAlaGlyLysProLeuValLysThrSerArgMetThrValIleAsn 1282
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
3614 TGTATGACAGGACCTGTCATCTGTCTGTGTGAAAGAAATATACCAAAA 3663
1283 ThrGluLysProAlaValThrValAspIleGlySerThrIleLysThrVa 1299
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
3664 CCAGAGCACAAACCATCTGCTCTGTGTGGAGCATCGTGGAGGCGAGC 3713
1299 lGlnGlyValAsnValThrIleAsnCysGlnValAlaGlyValProGluA 1316
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
3714 CCTTGGAGCAACGTCACATCCGATGTCCTGTAAAGGTGTCCTCAGC 3763
```

```
1316 laGluValThrTrpPheArgAsnLysSerLysLeu...GlySerProHis 1331
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
3764 CTAATATAACTTGGTTGAGAGAGGAGATCTCTGAGTGGCAATGTTTC 3813
1332 HisLeuHisGluGlySerLeuLeuThrAsnValSerSerSerAspG 1348
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
3814 TTCTTTTCAATGGATCCCTGTTGTCAGAAATGTTTCCCTTGAAAATGA 3863
1348 nGlyLeuTyrSerCysArgAlaAlaAsnLeuHisGlyGluLeuThrGluS 1365
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
3864 AGNACTACGTCTGCATAGCCACCAATGCTCTGGAAAGGACAGTGGCAA 3913
1365 erThrGlnLeuLeuLeuLeuAspProProGlnValProThrGlnLeuGlu 1381
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
3914 CATCTGTACTCCACTTCTGTGGAAGAGATGGCCAGAGAGTAGAATCGTA 3963
1382 AspIleArgAla.....LeuLeuAlaAlaThrGlyProAsnLe 1394
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
3964 TTCTGCAAGGACATATAAAGTACATCTCCAGGCACCAACACTAGAAC 4013
1394 uProSerValLeuThrSerProLeuGlyThrGlnLeuValLeuAspProG 1411
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
4014 CAACAGC.....AATGACCCA. 4029
1411 lyAsnSerAlaLeuLeuGlyCysProIleLysGlyHisProValProAsn 1427
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
4030 .....ACAGGAGAACCCCGCCT... 4047
1428 lIleThrTrpPheHisGlyGlyGlnProIleValThrAlaThrGlyLeuTh 1444
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
4047 ..... 4047
1444 rHisHisIleLeuAlaAlaGlyGlnIleLeuGlnValAlaAlaAsnLeuSerG 1461
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
4047 ..... 4047
1461 lyGlySerGlnGlyGluPheSerCysLeuAlaGlnAsnGluAlaGlyVal 1477
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
4047 ..... 4047
1478 LeuMetGlnLysAlaSerLeuValIleGlnAspTyrTrpTrpSerValas 1494
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
4048 .....CARGAGCCTTTTGGGACCTGG 4070
1494 pArgLeuAlaThrCysSerAlaSerCysGlyAsnArgGlyValGlnGlnP 1511
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
4071 TAACGTGTCACATTGTTCTGCCACCTGTGTCATTTGGGAGCCGCATTTC 4120
1511 roArgLeuArgCysLeuLeu...AsnSerThrGluValAlaAsnProAlaHis 1526
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
4121 ACAGACCCGACGTGTGTGTCGCAATGGCAGGAAGTAGTGAGGCCCTG 4170
1527 CysAlaGlyLysValArgProAlaValGlnProIleAlaCysAsnArgAr 1543
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
4171 TGTGATCACCTCCAGAGCCACTGGCTGGTTTGAGCCCTGTAACATCG 4220
1543 gAspCysProSerArgTrpMetValThrSerTrpSerAlaCysThrArgS 1560
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
4221 GGACTGCCAGCGAGGTGTTTCAAGTGTGTGTCACAGTGTCTGTGT 4270
1560 erCysGlyGlyValGlnThrArgArgValThrCysGlnLysLeuLys 1576
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
4271 CTTGGGTGAAGGATACCAAGTCGCGAGGTGACGTGCAAGCGGACAAAA 4320
1577 AlaSerGlyIleSerThrProValSerAsnAspMetCysThrGlnValAl 1593
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
4321 GCCAATGGACTGTCAGGTGGTGTCTCCAGAGCATGTGCCCT...AA 4367
1593 aLysArgProValAspThrGlnAlaCysAsnGlnGlnLeuCysValGluT 1610
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
4368 AGACCGGCTCTGGGAAGAAACCATGTTTGTGTCATCATCTGTGTTCAGT 4417
```


1268 GCTGAATGTGGATATCCGCTTGAGAGGGTAGTCTCTGACCAATTATG 1317
||||| |||:||||: ||:||||| |||:|||||
338 sh1stYrPrProGluAaSnileLysProLysProLysLeuGlnCysA 355
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1318 TCACACTACTACCCTGAAATGTAACCAACCAACCAACCAACCAACCA 1367
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
355 snLeuAspProCysProAlaSerAspGlyTyrLysGlnIleMetProTyr 371
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1368 GCATGATCCCTGCCCATCAAGTGTGATTTAAAGACATAATGCCCTAT 1417
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
372 AspLeuTyrHisProLeuProArg..TpGluAlaThrProTyrThrAla 387
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1418 GACCACTTCAACACCTCTCTCGAGCTGGGAACATAAATCCTTGGACTGCA 1467
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
388 CysSerSerSerCysGlyGlyGlyLeuGlnSerArgAlaValSerCysVa 404
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1468 TGTTCGCTGCTCTGGAGGAGGATTCAGAGACGAGCTTTGTGTGT 1517
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
404 lGluGlnAspIleGlnGlyHisValThrSerValGluGluTyrLysCysM 421
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1518 AGAGGAATCCATGCATCGAGAGATATTGCAGGTGGAAGAAATGGAAGTCCA 1567
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
421 etYrThrProLysMetProIleAlaGlnProCysAsnIlePheAspCys 437
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1568 TGTACGACCCCAACCCAGGTTATCAAACTGTAACTGTATTGTTGTC 1617
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
438 ProlysrProLeuAlaGlnGluThrSerProCysThrValThrCysGlyGl 454
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1618 CCCAAGTGGATTGCCATGGAGTGGTCTCAGTCACAGTGCATGTTGGCG 1667
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
454 nGlyLeuArgTyrArgValValLeuGlnCysIleAspHisArgGlyMetHisT 471
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1668 AGGGTTACGACTACCGGTTCTCTGTGTATTAACCAACCGCGAGAGCATG 1717
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
471 hrGlyGlyCysSerProLysThrLysProHisIleLysGluGluCysIle 487
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1718 TTGGGGCTGCATCCCACTCAACTGAGTTACACATCAAGAAGAAATGTGTC 1767
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
488 ValProThrProCysTyrLysProLysGluLysLeuProValGluAlaLyl 504
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1768 ATTCCTATCCCGTGTATTAACCAACCAACCAACCAACCAACCAACCA 1817
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
504 sLeuProThrPheLysGlnAlaGlnLeuGluGluGlyAlaAlaValS 521
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1818 ATTGCCCTGTGCTGAACCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAG 1867
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
521 erGluGluProSerPheIleProGluAlaThrSerAlaCysThrValThr 537
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1868 CAGAAAGCAACCACTTCATTCAGAACCCCTGGTCAGCCTGCAGTACCACG 1917
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
538 CysGlyValGlyThrGlnValArgIleValArgCysGlnValLeuLeuSe 554
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1918 TGTGGGCGAGGTGTGCAGTCCGCGAGGTGAAGTGGCGTGTGCTCTCATC 1967
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
554 rPheSerGlnSerValAlaAspLeuProIleAspGluCysGluGlyProL 571
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1968 ATTCAGCAGACTGAGACTGAGTGCCTCCGAGAGAGAGTGTCAAGGCCCA 2017
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
571 ysProAlaSerGlnArg.....AlacystYrAlaGlyPro 582
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
2018 AGTCTCCACCGACGGCCCTCCCTCCGAGGATGTGATGAGAGCCCG 2067
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
583 CysSerGlyGluIleProGluPheAsnProAspGluThrAspGlyLeuPh 599
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
2068 GCCTCCGAGAGCTAGACATCCCTCTCCCT..... 2097
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
599 eGlyGlyLeuGlnAspPheAspGluLeuTyrAspTyrPgluTyrGluGlyP 616
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
2098GAGGACAGTGAGACGACTTACGACTGGGAGTACCTGGGT 2137
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
616 heThrLysCysSerGluSerGlyGlyGlyValGlnGluAlaValVal 632
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||


```

||||: ||| :||||:||||: |||||:
4795 CCGAGAGGTGAGGATGCTTGGCGGCCAGCTGATATATACAGTGGTG 4844
1106 GlnGluIlePheArgSerHisLeuGluHisGlnAspThrLeuLeuLysPr 1122
||||: ||||||: ||| :|||
4845 GCCGAATTGACCAAGGACACCAACACATGCAG.....TGGCGGG 4888
1122 oSerGluArgThrSerProValThrLeuSerProHisLysHisVals 1139
||||: ||| |||
4889 CAUCCAGAGAGACACTCT..... 4910
1139 erGlyPheSerSerSerLeuArgThrSerSerThrGlyAspAlaGlyGly 1155
||||: ||||| |||:||||: |||:
4911 .....GCTGCTCAGCTCAGA.....GGGAAACAGGGAGT 4940
1156 GlySerArgProHisArgLys..... 1163
||||: ||| |||
4941 GTGCTCCAAAGCTCGCATGCAAAACACTGCGCAAGCTGACATTCAGGCC 4990
1164 .....ProThrIleLeuArgLysIleSerAlaAlaGlnGlnLeuSerA 1178
||||: |||:||||: ||| |||
4991 GAAGGACCTGTTCTCATGAGCAA...AGCCAACCTCCCTCAATTTCAT 5037
1178 laSerGluValVal...ThrHisLeuGlyGlnThrValAlaLeuAlaSer 1193
||||: |||:||||: ||| |||
5038 TTAATAAACAAATAAATCCAGGATTGAAATACAGTATACATTACAAAA 5087
1194 GlyThrLeuSerValLeuLeuHisCysGluAlaIleGlyHisProArgPr 1210
||||: |||: ||| |||
5088 AGGACAGAGGTCATCAATACTGTGTGACCTTATTACCCCGAGTGAGGC 5137
1210 oThrIleSerTrpAlaArgAsnGlyGluValGlnPheSerAspArgI 1227
||||: |||: ||| ||| ||| |||
5138 CACATATACATGGACCAAGATGGAACTGTGTACAGCCCTCAGTAAAAA 5187
1227 leLeuLeuGlnProAspAspSerLeuGlnIleLeuAlaProValGluAla 1243
||||: |||: ||| ||| ||| |||
5188 TAATTTGGATGGAACCTGGGAAGATACAGATACAGAATCCTACAAAGAAA 5237
1244 AspValGlyPheThrCysAsnAlaThrAsnAlaLeuGlyTyTrAspSe 1260
||||: |||: ||| ||| ||| ||| |||
5238 GAACNAGCATATATGAATGTCTGTACTATCATCTGTGTACAGATGT 5287
1260 rValSerIleAlaValThrLeuAlaGlyLysProLeuValLysThrSerA 1277
||||: |||: ||| ||| ||| ||| |||
5288 GGAAGATTCTCTGTGCTGTATGCGAGGACACCTGTCTATCTGTCTGTG 5337
1277 rgMetThrValIleAsnThrGluLysProAlaValThrValAspIleGly 1293
||||: |||: ||| ||| ||| ||| |||
5338 AAAGAAATATATCAAAACAGACCAACCATCTCTGTCTGTGTGTGGA 5387
1294 SerThrIleLysThrValGlnGlyValAsnValThrIleAsnCysGlnVa 1310
5388 GCATCGTGGAGGAGCCCTTGGAGCAACGTGACAAATCCGATGCTCTGT 5437
1310 laGlyValProGluAlaGluValThrTrpPheArgAsnLysSerLysL 1327
||||: |||: ||| ||| ||| ||| ||| |||
5438 AAAAGGTGTCCTCAGCCTAATATACTTGGTTGAAGAGAGGATCTC 5487
1327 eu...GlySerProHisHisLeuHisGluGlySerLeuLeuLeuThrAsn 1342
||||: |||: ||| ||| ||| ||| ||| ||| |||
5488 TGAGTGGCAATTTCTTCTTTCATGTGATCCCTGTGTGTGTGAGAT 5537
1343 ValSerSerAspGlnGlyLeuTyTrSerCysArgAlaAlaAsnLeuHi 1359
||||: |||: ||| ||| ||| ||| ||| ||| |||
5538 GTTTCCTTGAATATGAAGAACCTACGCTGCATAGCAACCAATGCTCT 5587
1359 sgGlyLeuLeuThrGluSerThrGlnLeuLeuLeuLeuAspProGlnV 1376
||||: |||: ||| ||| ||| ||| ||| ||| |||
5588 TGGAAAGGAGTGGCAACATCTGTACTCCACTTGTGCGAACGAAGATGC 5637
1376 aIProThrGlnLeuGluAspIleArgAla.....LeuLeuAla 1388
||||: |||: ||| ||| ||| ||| ||| ||| |||

```

```

5638 CAGAGAGTAGAATCGTATTTCTGCAGGACATAAAAGTACATTCCTCCAG 5687
1389 AlaThrGlyProAsnLeuProSerValLeuThrSerProLeuGlyThrGl 1405
||||: |||: ||| |||
5688 GCAACCACTAGAACCAACAGC..... 5711
1405 nLeuValLeuAspProGlyAsnSerAlaLeuLeuGlyCysProIleLysG 1422
||||: ||| |||
5712 .....AATGACCCA.....ACAG 5724
1422 lyHisProValProAsnIleThrTrpPheHisGlyGlyGlnProIleVal 1438
||||: ||| ||| |||
5725 GAGAACCCCGCCT..... 5738
1439 ThrAlaThrGlyLeuThrHisHisIleLeuAlaAlaGlyGlnIleLeuGl 1455
5738 ..... 5738
1455 nValAlaAsnLeuSerGlyGlySerGlnGlyGluPheSerCysLeuAlaG 1472
5738 ..... 5738
1472 InAsnGluAlaGlyValLeuMetGlnLysAlaSerLeuValIleGlnAsp 1488
||||: |||: ||| |||
5739 .....CAAGAG 5744
1489 TyrTrpTrpSerValAspArgLeuAlaThrCysSerAlaSerCysGlyAs 1505
||||: |||: ||| ||| ||| ||| |||
5745 COTTTTGGGAGCCTGGTAACCTGGTCACATTGTTGTCACCTGTGTGCA 5794
1505 nArgGlyValGlnGlnProArgLeuArgCysLeuLeu...AsnSerThrG 1521
||||: |||: ||| ||| ||| ||| ||| |||
5795 TTTGGAGCGCCCATTCAGAGACCCAGTGTGTGATGGCCAAATGGCAGG 5844
1521 luValAsnProAlaHisCysAlaGlyLysValArgProAlaValGlnPro 1537
||||: |||: ||| ||| ||| ||| ||| |||
5845 AAGTAGTGAGGCCCTGTGTATCACCTCCAGAGCCACTGGCTGGGTTT 5894
1538 IleAlaCysAsnArgAspCysProSerArgTrpMetValThrSerTr 1554
||||: |||: ||| ||| ||| ||| ||| ||| |||
5895 GAGCCTGTAAACATCGGCACATCCAGCAGGTGTTTCCACAGTGTG 5944
1554 pSerAlaCysThrArgSerCysGlyGlyValGlnThrArgArgValT 1571
||||: |||: ||| ||| ||| ||| ||| ||| |||
5945 GTACAGTGTCTGTGTCTGTGCGTGAAGATACACAGTCGCGAGTGA 5994
1571 hrCysGlnLysLeuLysAlaSerGlyIleSerThrProValSerAsnAsp 1587
||||: |||: ||| ||| ||| ||| ||| ||| |||
5995 CGTGCAAGCGGACAAAGCAATGGAACGTGCAGGTGGTGTCTCCAAG 6044
1588 MetCysThrGlnValAlaLysArgProValAspThrGlnAlaCysAsnGl 1604
||||: |||: ||| ||| ||| ||| ||| ||| |||
6045 GCAITGCCCCCT...AAGACCGCCCTCTGGGAAGAAACCACTGTTTGG 6091
1604 nGlnLeuCysValGluTrpAlaPheSerSerTrpGlyGlnCysAsnGlyP 1621
||||: |||: ||| ||| ||| ||| ||| ||| ||| |||
6092 TCATCATGTGTTCAGTGG.....GAACAGGAACCGGTGTCTCTGGAC 6135
1621 roCysIleGlyProHisLeuAlaValGlnHisArgGlnValPheCysGln 1637
||||: |||: ||| ||| ||| ||| ||| ||| ||| |||
6136 GTTGATGGCCCTGCTGTGAGGATGCAGCAGGTGCACAGCTTGTCAA 6185
1638 ThrArgAspGlyIleThrLeuProSerGluGlnCysSerAlaLeuProAr 1654
||||: |||: ||| ||| ||| ||| ||| ||| ||| |||
6186 CACAACAGCTCT.....GACTCAACTGTGATCAGCAAAAGAG 6223
1654 gProValSerThrGlnAsnCysTrpSerGluAlaCysSerValHisTrpA 1671
||||: |||: ||| ||| ||| ||| ||| ||| ||| |||
6224 ACCCACTTAAGAGGAGACATGCATACGGGCTGTGATGTGTGTGGC 6273
1671 rgValSerLeuThrThrLeuCysThrAlaThrCysGlyAsnTyrglyPhe 1687
||||: |||: ||| ||| ||| ||| ||| ||| ||| |||
6274 ACAGAGCCCTTGGAAAGCCCTGTACAGCAGCTGTGGCAGG...GGTTTC 6320

```



```
234 lYhrLysGlyGluAnSerLeuSerSerThrGlyThrPheLeuValAsp 250
|||||
751 GGACTAAAGGTGAAACAGTCTCAGCTCCACAGAACTTTCTTGTGGAC 800
|||||
251 AsnSerSerValAspPheGlnLysPheProAspLysGluLeuLeuArgMe 267
|||||
801 AATTCTAGTGTGGACTTCCAGAAATTTCCAGACAAAGAGATACTGAGAA 850
|||||
267 tAlaGlyProLeuThrAlaAspPheIleValLysIleArgAsnSerGlyS 284
|||||
851 GGCTGGACCACTCACACAGATTTCATTGTCAGAGATTCTGTAATCTGGGCT 900
|||||
284 erAlaAspSerThrValGlnPheIlePheTyrglnProIleIleHisArg 300
|||||
901 CCGCTGACAGTACAGTCCAGTTCATCTTCATCAACCCATCATCCACCGA 950
|||||
301 TrpArgGluThrAspPhePheProCysSerAlaThrCysGlyGlyTy 317
|||||
951 TGGAGGAGAGCGGATTTCTTTCTTGTCTCAGCAACCTGTGGAGGAGGTTA 1000
|||||
317 rGlnLeuThrSerAlaGluCysTyAspLeuArgSerAsnArgValVala 334
|||||
1001 TCACCTGACATCGCTCAGTCTACGATCTCAGGAGCAACCGTGTGGTGG 1050
|||||
334 laAspGlnTyrCysHisTyThrProGluAsnIleLysProLysProLys 350
|||||
1051 CTGACCAATAGTCTACATATTACCCAGAGAAACATCAAAACCCCAACG 1100
|||||
351 LeuGlnGluCysAsnLeuAspProCysProAlaSerAspGlyTyrLysG 367
|||||
1101 CTTCAGGAGTGCAACTTGGATTCCTGTGTCAGCCAGTACGGATACAGCA 1150
|||||
367 nIleMetProTyrAspLeuTyrHisProLeuProArgTyrGluAlaThrP 384
|||||
1151 GATCATGCCATTATCACCTTACCATCCCTCTCTCGTGGAGGCCACCC 1200
|||||
384 roTTrpAlaCysSerSerCysGlyGlyGlyIleGlnSerArgAla 400
|||||
1201 CATGGACCGCTGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1250
|||||
401 ValSerCysValGluGluAspIleGlnHisValThrSerValGluGlu 417
|||||
1251 GTTTCCTGTGTGGAGGAGGAGCATCCAGGGGATGTACCTTCAGTGGAA 1300
|||||
417 uTrpLysCysMetTyrThrProLysMetProIleAlaGlnProCysAsnI 434
|||||
1301 GTGGAATGTCATGTACACCCCTAAGATGCCCATCGCGCAGCCCTGCAACA 1350
|||||
434 lePheAspCysProLysTrpLeuAlaGlnGluTrpSerProCysThrVal 450
|||||
1351 TTTTGTAGTCCCTTAAATGGCTGGCACAGGAGTGGTCTCCGTGCACAGT 1400
|||||
451 ThrCysGlyGlnGlyLeuArgTyrArgValValLeuCysIleAspHisAr 467
|||||
1401 ACGTGTGGCCAGGCGCTCAGATACCGTGTGGTCTCTCTCATCGACCATCG 1450
|||||
467 gGlyMetHisThrGlyGlyCysSerProLysThrLysProHisIleLysG 484
|||||
1451 AGGAATGCACACAGGAGCGCTGTAGCCCAAAACAAAGCCCAACATAAAG 1500
|||||
484 luGluCysIleValProThrProCysTyrLysProLysGluLysLeuPro 500
|||||
1501 AGGAATGATCGTACCCATCCCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1550
|||||
501 ValGluAlaLysLeuProTyrPheLysGlnAlaGlnGluLeuGluGluG 517
|||||
1551 GTCCAGGCGCAAGTTCGCAATGTTCAACAGCTCAAGAGCTAGAGAAGG 1600
|||||
517 yAlaAlaValSerGluProProPheIleProGluAlaTrp..... 531
|||||
1601 ACCTGCTGTGTAGAGGAGCCCTCGTAAGTGTAAAGACACAGACTGTTC 1650
|||||
532 .....SerAlaCysThrValThrCys 538
```

```
1651 TATATTGAAACTTTTGTTTAAAGAAAGCAGTGTCTCACTGGTGTAGCT 1700
|||||
539 GlyValGlyThrGlnValArgIleVal 547
|||||
1701 TTCATGGTCTGAACTAAGTGTATTC 1727
|||||
seq_name: gb_pat:AX224819
seq_documentation_block:
LOCUS AX224819 2538 bp DNA linear PAT 10-SEP-2001
DEFINITION Sequence 11 from Patent WO0161011.
ACCESSION AX224819
VERSION AX224819.1 GI:15554914
KEYWORDS
SOURCE
human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
1 (bases 1 to 2538)
AUTHORS Donoho,G., Scoville,J., Turner,C.A., Friedrich,G., Zambrowicz,B.
and Sands,A.T.
TITLE Novel human thrombospondin repeat proteins and polynucleotides
encoding the same
JOURNAL Patent: WO 0161011-A 11 23-AUG-2001;
Lexicon Genetics Incorporated (US)
FEATURES
Location/Qualifiers
source 1..2538
/organism="Homo sapiens"
/db_xref="taxon:9606"
BASE COUNT 670 a 627 c 678 g 563 t
ORIGIN
alignment_scores:
Quality: 2809.50 Length: 783
Ratio: 4.302 Gaps: 4
Percent Similarity: 83.397 Percent Identity: 60.920
alignment_block:
US-10-044-807-2 x AX224819 ..
Align seg 1/1 to: AX224819 from: 1 to: 2538
22 SerSerArgThrAlaArgSerGluGluAspArgAspGlyLeuTrpAspAl 38
|||||
190 ACCTCAAGAAACACTCGTTTCAGATGAAGAAAGATGGCACTGGGATGC 239
|||||
38 aTrpGlyProTrpSerGluCysSerArgThrCysGlyGlyAlaSerT 55
|||||
240 TTGGGGCGACTGGAGTGACTGCTCCCGGACCTGTGGGGGAGGAGGATCAT 289
|||||
55 yrSerLeuArgArgCysLeuSerSerLysSerCysGluGlyArgAsnIle 71
|||||
290 ATTCCTCGGAGATGTTGACTGGAGGAATGTGAAGGGCAGACATTT 339
|||||
72 ArgTyrArgThrCysSerAsnValAspCysProProGluAlaGlyAspPh 88
|||||
340 CGGTACAAGACATGCGCAATCATGACTGCCCTCCAGATGCAAGATTT 389
|||||
88 eaArgAlaGlnGlnCysSerAlaHisAsnAspValLysHisGlyGlnP 105
|||||
390 CAGAGCCCGCAGCAGTGTCTCAGCCTACAAATGATGTCCAGTATCAGGGCA 439
|||||
105 heTyrGluTrpLeuProValSerAsnAspProAspAsnProCysSerLeu 121
|||||
440 ACTATGAATGGCTTCCACGATATATGATGCTGCTGCCCGGTGTGCATC 489
|||||
122 LysCysGlnAlaLysGlyThrThrLeuValValGluLeuAlaProLysVa 138
|||||
490 AAGTGTTCATGCACAGGACAAACACTTGGTGGAGCTGGCACCTAAGGT 539
|||||
138 lLeuAspGlyThrArgCysTyrThrGluSerLeuAspMetCysIleSerG 155
```

|||||
540 ACTGATGGAACCTCGTTGCAACACGAGCTCCCTGGACATGTGTATCAGTG 589
155 lYleuCysGlnIleValGlyCysAspHisGlnLeuGlySerThrVallys 171
|||||
590 GCATCTGTCCAGCAGCTGGCTCGATCGCACTGGGAACCAATGCCAAG 639
172 GluAspAsnCysGlyValCysAsnGlyAspGlySerThrCysArgLeuVa 188
|||||
640 GAGGACAACTGTGGAGTCTGTCCGCGGATGCTCCACCTGCAGGCTGT 689
188 lArgGlyGlnTyrLysSerGlnLeuSerAlaThrLysSerAspAspThrV 205
|||||
690 ACGGGACAACTCAAGTCCACAGCTTTCCTGAAAAAAGAGAAATG 739
205 alValAlaIleProTyrGlySerArgHisIleArgLeuValLeuLysGly 221
|||||
740 TAATTGCTGTTCTTGGGAAGTCGAAGTCAGAAATTACAGTGAAGGA 789
222 ProAspHisLeuTyrLeuGluThrLysThrLeuGlnGlyThrLysGlyG 238
|||||
790 CCTGCCCACTCTTATTGAAATCAAAAACACTTCAAGGAAGCAAGGAGA 839
238 uAsnSerLeuSerSerThrGlyThrPheLeuValAspAsnSerSerValA 255
|||||
840 ACACAGCTTTACAGCCCGCGCTTTGCTGAGAAAACCAACAGTGG 889
255 spPheGlnLysPheProAspLysGluIleLeuArgMetAlaGlyProLeu 271
|||||
890 AATTTCAGAGGGGCTCCGAGAGGCAAACTTTAAGATTCCAGGACCTCTG 939
272 ThrAlaAspPheIleValLysIleArgAsnSerGlySerAlaAspSerTh 288
|||||
940 ATGCTGATTTCACTTCAAGACAGGCTACACTGCAGCAAGACAGCGT 989
288 rValGlnPheIlePheTyrGlnProIleIleHisArgTrpArgGluThra 305
|||||
990 GGTTCAGTCTCTTTACAGCCCATCAGTCATCAGTGGAGACAAACTG 1039
305 spPhePheProCysSerAlaThrCysGlyGlyTyrGlnLeuThrSer 321
|||||
1040 ACTTCTTCCCTGCACTGTGACGTGTGGAGGAGTTATCAGTCAATTC 1089
322 AlaGlyCysTyrAspLeuArgSerAsnArgValValAlaAspGlnTyrCy 338
|||||
1090 GCTGAATGTGGATATCCGCTGAAGAGGGTAGTCTCTGACCATATTG 1139
338 sHisTyrTyrProGluAsnIleLysProLysProLysLeuGlnCysA 355
|||||
1140 TCACCTACTACCTGAAAAATGTAAACCAAAACCAAACTGAAGAAATGCA 1189
355 snLeuAspProCysProAlaSerAspGlyTyrLysGlnIleMetProTyr 371
|||||
1190 GCATGGATCCTGCCCATCAAGTATGATGGATTAAAGAGATAATGCCCTAT 1239
372 AspLeuTyrHisProLeuProArgTrpGluAlaThrProTrpThrAlaCy 388
|||||
1240 GACCACCTCAACCTCTCTCTCGCTGGACACATAATCCTTGGACTGCATG 1289
388 sSerSerSerCysGlyGlyIleGlnSerArgAlaValSerCysValG 405
|||||
1290 TTCGTGCTCTGTGGAGGAGGATTCAGACACGGAGCTTGTGTCTGTAG 1339
405 luGluAspIleGlnGlyHisValThrSerValGluGluTrpLysCysMet 421
|||||
1340 AGGAATCCATGCATGGAGAGATATTGACGTTGGAAGAAATGGAATGCATG 1389
422 TyrThrProLysMetProIleAlaGlnProCysAsnIlePheAspCysPr 438
|||||
1390 TAGCACCCCAACCCCAAGGTTTGAACAACTGTAATCTGTTGATTGCC 1439
438 oLysTrpLeuAlaGlnGluTrpSerProCysThrValThrCysGlyGlnG 455
|||||

1440 CAAGTGAATTGCCATGGAGTGGTCTTCAGTGCACAGTACTTGTGGCGGAG 1489
455 lYleuArgTyrArgValValLeuCysIleAspHisArgGlyMetHisThr 471
|||||
1490 GGTACGGTACCGGTTGTCTGTGTATTAAACCAACCGCGAGAGCATGTT 1539
472 GlyGlyCysSerProLysThrLysProHisIleLysGluGluCysIleVa 488
|||||
1540 GGGGCTGCATCCACACTGAAGTTACATCAATCAAGAAGAATGTGTCTAT 1589
488 lProThrProCysTyrLysProLysGluLysLeuProValGluAlaLysL 505
|||||
1590 TCCCATCCCGTGTATATAAACCAAAAGAAAAAGTCCAGTGGAAAGCAAAAT 1639
505 euProTrpPheLysGlnAlaGlnLeuGluGluGluGlyAlaLaValSer 521
|||||
1640 TGCCCTGGCTGAAACAAGCACAAGAACTAGAAGAGAGACCAGAAATAGCA 1689
522 GluGluProSerPheIleProGluAlaTrpSerAlaCysThrValThrCy 538
|||||
1690 GAAGAACCACCTTCATTCAGAACCTGTGTGAGCTGCAGTACCACGTG 1739
538 sGlyValGlyThrGlnValArgIleValArgCysGlnValLeuLeuSerP 555
|||||
1740 TGGGCGAGGTGTCAGTCCGCGAGTGAAGTCCGCTGTGCTCCTCACAT 1789
555 heSerGlnSerValAlaAspLeuProIleAspGluCysGluGlyProLys 571
|||||
1790 TCACCGAGACTGAGACTGAGTGCCTGCGGAGGAGAGTGTGAAGGCCCAAG 1839
572 ProAlaSerGlnArg.....AlaCysTyrAlaGlyProCy 583
|||||
1840 CTGCCACCGAAGCGCTCCCTCTCTGGNAGCATGTGATGAGAGCCCGC 1889
583 sSerGlyGluIleProGluPheAsnProAspGluThrAspGlyLeuPheG 600
|||||
1890 CTCCGAGAGCTAGACATCCCTCTCCCT..... 1917
600 lYglyLeuGluAspPheAspGluLeuTyrAspTrpGluTyrGluGlyPhe 616
|||||
1918GAGGACAGTGGAGAGCTTACGACTGGGAGTACGCTGGGTTC 1959
617 ThrLysCysSerGluSerCysGlyGlyValGlnGluAlaValValSe 633
|||||
1960 ACCCTTCGACAGCAACATGCTGGAGGCCATCAAGAACCTATGACAGT 2009
633 rCysLeuAsnLysGlnThrArgGluProAlaGluGluAsnLeuCysValT 650
|||||
2010 CTGCTTACATATCCAGACCCAGACACAGTCAATGACAGCTTGTGTGATA 2059
650 hrSerArgArgProGlnLeuLysSerCysAsnLeuAspProCys 666
|||||
2060 TGTTCACCGTCTCCAGCATGAGCCAGGCTGTATACACAGAGCCCTGT 2109
667 ProAlaArgTrpGluIleGlyLysTrpSerProCysSerLeuThrCysGl 683
|||||
2110 CCCCCCAGGTGGCATGTGGGCTCTTGGGGCCCTGCTCAGCTACCTGTGG 2159
683 yValGlyLeuGlnThrArgAspValPheCysSerHisLeuLeuSerArgG 700
|||||
2160 AGTTGGAATTCAGACCCGAGATGTACTTGCCTGCAC..... 2196
700 luMetAsnGluThrValIleLeuAlaAspGluLeuCysArgGlnProLys 716
|||||
2197 ..CCAGGGGAGACCCCTGCCCTCCTCCTGAGGAG...TGCCGAGATGAAGAAG 2241
717 ProSerThrValGlnAlaCysAsnArgPheAsnCysProProAlaTrpTy 733
|||||
2242 CCCCATGCTTACAAAGCATGCAATCAGTTGACTGCCCCCTCTCTGCTGGCA 2291
733 rProAlaGlnTrpGlnProCysSerArgThrCysGlyGlyValGlnL 750
|||||
2292 CATTGAAGAATGGCAGCAGTGTCCAGGACTGTGCGGGGGAACCTCAGA 2341

```
alignment_scores:
  Quality: 2625.50
  Ratio: 4.361
```

```

1214 GGTTCAGTTCTCTTTTACAGCCCATCAGTCATCAGTGGAGACAAACTG 1263
305 spPheProCysSerAlaThrCysGlyGlyTyrGlnLeuThrSer 321
1264 ACTTCTTTCCCTGCACTGAGCTGAGTGGAGGAGTTATCAGCTCAATTCT 1313
322 AlaGluCysTyAspLeuArgSerAsnArgValValAlaAspGlnTyrCy 338
1314 GCTGAATGCTGGATATCCGCTGAAGAGGGTAGTTCCTGACCAATTATG 1363
338 shiTyTyTyProGluAsnIleLysProLysProLysLeuGlnGluCysA 355
1364 TCACTACTACCTGAAATGTAACCAAAACCAAACTGAAGGAATGCA 1413
355 snLeuAspProCysProAlaSerAspGlyTyrLysGlnIleMetProTyr 371
1414 GCATGATCCCTGCCCATCAAGTATGATGATTTAAAGAGATAATGCCCTAT 1463
372 AspLeuTyrHisProLeuProArgTrpGluAlaThrProTrpThrAlaCy 388
1464 GACCATTCCCAACCTTCTCTCGCTGGGAACATAATCCTTGGACTGCATG 1513
388 sSerSerCysGlyGlyGlyGlnSerArgAlaValSerCysValG 405
1514 TTCCGTGCTCTGGAGAGGAGGATTCAGAGCGGAGCTTGTGTGTGTAG 1563
405 luGluAspIleGlnGlyHisValThrSerValGluGluTrpLysCysMet 421
1564 AGGAATCATCATGATGAGAGATATTGCAAGTGGAAAGATGGAAGTGCATG 1613
422 TyrThrProLysMetProIleAlaGlnProCysAsnIlePheAspCysPr 438
1614 TAGCCACCAACCAACCAAGTTATGCAAACTTGTAATCTGTTGATTGCC 1663
438 olYstTrpLeuAlaGlnGluTrpSerProCysThrValThrCysGlyGlnG 455
1664 CAAGTGGATTGCGATGGAGTGGTCTCAGTGCACAGTACTTGTGGCCGAG 1713
455 lyLeuArgTyrArgValValLeuCysIleAspHisArgGlyMetHisThr 471
1714 GGTACGGTACCGGGTGTCTGTGTATTAACCAACCGGAGAGCATGTT 1763
472 GlyGlyCysSerProLysThrLysProHisIleLysGluGluCysIleVa 488
1764 GGGGCTGCATCCCACTCAAGTTACATCAAGAAAGAAATGTGTGCAT 1813
488 lProThrProCysTyrLysProLysGluLysLeuProValGluAlaLysL 505
1814 TCCATCCCTGTTATACCAACCAAGAAAGTCCAGTGGAGCAAAAT 1863
505 euProTrpPheLysGlnAlaGlnGluLeuGluGlyAlaAlaValSer 521
1864 TGCCTTGGCTGAAACAGCAAGCAAGCAATAGAACAGACCAATAGCAACA 1913
522 GluGluProSerPheIleProGluAlaTrpSerAlaCysThrValThrCy 538
1914 GAAGAACCAACGTTTCATCCAGAACCTGGTCACCCCTGCATCCACAGTG 1963
538 sGlyValGlyThrGlnValArgIleValArgCysGlnValLeuLeuSerP 555
1964 TGGGCCGGGTGTGCGAGTCCGCGAGGTGAAGTCCGCTGCTCTCCAT 2013
555 heSerGlnSerValAlaAspLeuProIleAspGluCysGluGlyProLys 571
2014 TCAGCGAGACTGAGACTGAGTCCCGGAGGAAGAGTGTGAAGGCCCAAG 2063
572 ProAlaSerGlnArg.....AlaCysTyrAlaGlyProCy 583
2064 CTGCCCCACCAAGGCCCTGCTCTCGAGAGCATGTGATGAGAGCCCGC 2113
583 sSerGlyGluIleProGluPheAsnProAspGluThrAspGlyLeuPheG 600
2114 CTCGCGAGACTAGACATCCCTCTCCCT..... 2141

```

```

600 lyGlyLeuGlnAspPheAspGluLeuTyrAspTrpGluTyrGluGlyPhe 616
2142 .....GAGGACAGTGGAGACGACTTACGACTGGGAGTACGGTGGTTC 2183
617 ThrLysCysSerGluSerCysGlyGlyValGlnGluAlaValValse 633
2184 ACCCTTGCACAGCAACATCTTGGGAGGCCATCAGAAGCCATAGCAGT 2233
633 rCysLeuAsnLysGlnThrArgGluProAlaGluGluAsnLeuCysValT 650
2234 GTGCTTACATATCCAGACCAGCAGACAGTCAATGACAGCTTGTGTGATA 2283
650 hrSerArgArgProGlnLeuLeuLysSerCysAsnLeuAspProCys 666
2284 TGTTCACCGCTCTCCAGCCATGAGCCAGGCTGTAAACAGAGCCCTGT 2333
667 ProAlaArgTrpGluIleGlyLysTrpSerProCysSerLeuThrCysG1 683
2334 CCCCCAGGTGGCATGTGGCTCTTGGGGCCCTGCTACGCTACCTGTGTG 2383
683 yValGlyLeuGlnThrArgAspValPheCysSerHisLeuLeuSerArg 700
2384 AGTTGAATTACAGCCGAGATGTGTACTGCCCTGCAC..... 2420
700 luMetAsnGluThrValIleLeuAlaAspGluLeuCysArgGlnProLys 716
2421 ..CCAGGGAGAGCCCTGCCCTCTCTGAGGAG...TGCCGAGATGAAAG 2465
717 ProSerThrValGlnAlaCysAsnArgPheAsnCysProProAlaTrpTy 733
2466 CCCCATGCTTTACAGCATGCAATCAGTTTGAATGCCCTCTCTGCGCA 2515
733 r 733
2516 c 2516

```

seq_name: gb_pat:AX224821

seq_documentation_block:

LOCUS AX224821 2316 bp DNA linear PAT 10-SEP-2001
 DEFINITION Sequence 13 from Patent WO0161011.
 ACCESSION AX224821
 VERSION AX224821.1 GI:15554915
 KEYWORDS

SOURCE human.

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Euthera; Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1 (bases 1 to 2316)
 AUTHORS Donoho,G., Scoville,J., Turner,C.A., Friedrich,G., Zambrowicz,B.
 and Sands,A.T.

TITLE Novel human thrombospondin repeat proteins and polynucleotides
 encoding the same

JOURNAL Patent: WO 0161011-A 13 23-AUG-2001;

Lexicon Genetics Incorporated (US)

FEATURES Location/Qualifiers

1..2316

/organism="Homo sapiens"

/db_xref="taxon:9606"

BASE COUNT 617 a 575 c 616 g 508 t

ORIGIN

alignment_scores:

Quality: 2615.50 Length: 741

Ratio: 4.267 Gaps: 4

Percent Similarity: 82.726 Percent Identity: 60.054

alignment_block:

US-10-044-807-2 x AX224821 ..

Align seg 1/1 to: AX224821 from: 1 to: 2316

364 GlyTyrLysGlnIleMetProTyrAspLeuTyrHisProLeuProArgTyr 380
 994 GGATTTAAAGAGATATATGCCCTATGACCACTTCCCAACCTCTTCCTCGGTG 1043
 380 pGluAlaThrProTyrThrAlaCysSerSerCysGlyGlyGlyLeu 397
 1044 GGAACATAATCCTTGGACTGCATGTTCCGTGTCCTGTGGAGGAGATTC 1093
 397 InSerArgAlaValSerCysValGluGluAspIleGlnGlyHisValThr 413
 1094 AGACGACGGAGCTTGTGTGTGTAGAGGAATCCATGCATGGAGAGATATG 1143
 414 SerValGluGluTyrPylCysMetIleThrProLysMetProIleAlaGln 430
 1144 CAGGTGGGAAGATGAAGTGCATGTACGCCAACCCCAACCCCAAGGTATGCA 1193
 430 nProCysAsnIlePheAspCysProLysTriPLeuAlaGlnGluTyrSerP 447
 1194 AACTTGTATCTGTTGATGCCCCCAAGTGATGCCATGGAGTGGTCTC 124
 447 roCysThrValThrCysGlyGlnGlyLeuArgTyrArgValValLeuCys 463
 1244 AGTGACACAGTACTTGTGGCCGAGGGTTACGGTACCGGGTTTCTGTGT 1293
 464 IleAspHisArgGlyMethIleThrGlyGlyCysSerProLysThrLysP 480
 1294 ATTACACCACCGGAGAGCATGTGTGGGGCTGCAATCCACACTGAAGTT 1343
 480 ohIleLysGluGluCysIleValProThrProCysTyrLysProLysG 497
 1344 ACACATCAAGAAGAAATGTGTATTCCTCCATCCCGTCTTATAAACCAAAAG 1393
 497 LuLysLeuProValGluAlaLysLeuProTyrPheLysGlnAlaGlnGlu 513
 1394 AAAAAAGTCCAGTGGAGCAAAATTCCTTGGCTGAACACAGCACAGAA 1443
 514 LeuGluGluGlyAlaAlaValSerGluGluProSerPheIleProGluAl 530
 1444 CTAGAAGAGACCAAGATAGCACAGAGAACAACGTTCTATCCAGAAC 1493
 530 atrpSerAlaCysThrValThrCysGlyValGlyThrGlnValArgIleV 547
 1494 CTGGTCAGCCTGCAGTACCACGTGTGGCCAGGTGTGCAGGTCCGCGAGG 1543
 547 alArgCysGlnValLeuLeuSerPheSerGlnSerValAlaAspLeuPro 563
 1544 TGAAGTCCGTGTCTCTACATTCACGACAGACTGAGACTGAGCTGCC 1593
 564 IleAspGluCysGluGlyProLysProAlaSerGlnArg 576
 1594 GAGGAAGAGTGTGAAGGCCCCCAAGCTGCCACCGAAGCGCCCTGCCCTCT 1643
 577AlaCysTyrAlaGlyProCysSerGlyGluIleProGluPheAsn 592
 1644 GGAAGCATGTGACAGAGCCCGCCCTCCCGAGACTAGACATCCCTCTCC 1693
 592 roAspGluThrAspGlyLeuPheGlyLeuGlnAspPheAspGluLeu 608
 1694 CT.....GAGGACAGTGGAGCGACT 1713
 609 TyrAspTyrPLeuTyrGluGlyPheThrLysCysSerGluSerGlyGly 625
 1714 TACGACTGGAGTACGCTGGGTTCACCCCTTGACACACACATGCTGGG 1763
 625 yGlyValGlnGluAlaValSerCysLeuAsnLysGlnThrArgGluP 642
 1764 AGGCCATCAAGAACCCATAGCAGTGTCTTACATATCCAGACCCAGAGA 1813
 642 roAlaGluGluAsnLeuCysValThrSerArgArgProGluLeuLeu 658
 1814 CAGTCAATGACAGCTTGTGTGATATGGTCCACCGTCTCTCCAGCCATGAGC 1863
 659 LysSerCysAsnLeuAspProCysProAlaArgTyrGluIleGlyLysT 675

```

1864 CAGGCTGTAAACACAGAGCCCTGTCCCCAGTGGCATGTGGCTCTTG 1913
      :::::||||| :::::|||||  |||||:::|||||:::|||||
675 pSerProCysSerLeuThrCysGlyValGlyLeuGlnThrArgAspValp 692
      |||||:::|||||  |||||:::|||||:::|||||:::|||||
1914 GGGGCTCTCTCAGCTACCTGTGGAGTGGGAATTCAGACCCGAGATGT 1963
      ::|||::|  |||||:::|||||:::|||||:::|||||
692 heCysSerHisLeuLeuSerArgGluMetAsnGluThrValIleLeuAla 708
      ::|||::|  |||||:::|||||:::|||||:::|||||
1964 ACTGCTGCAC.....CCAGGGGAGACCCCTCCCTCCT 1998
      ::|||::|  |||||:::|||||:::|||||:::|||||
709 AspGluLeuCysArgGlnProLysProSerThrValGlnAlaCysAsnAr 725
      ::|||::|  |||||:::|||||:::|||||:::|||||
1999 GAGGAG...TGCGAGATGAAGAGCCCATGCTTACAGCATGCAATCA 2045
      ::|||::|  |||||:::|||||:::|||||:::|||||
725 gPheAsnCysProProAlaTrpTrpProAlaGlnTrpGlnProCysSera 742
      ::|||::|  |||||:::|||||:::|||||:::|||||
2046 GTTTCAGTGCCTCTGCTGGCAGCATTAAGAAATGGCAGCATGTCTCA 2095
      ::|||::|  |||||:::|||||:::|||||:::|||||
742 rgThrCysGlyGlyValGlnLysArgGluValLeuLeuCysLysGlnArg 758
      |||||:::|||||  |||||:::|||||:::|||||:::|||||
2096 GGACTGTGGGGGGAAGTCTCAGAAACAGAGAGTCACTGTGGCAGCTG 2145
      ::|||::|  |||||:::|||||:::|||||:::|||||
759 MetAlaAspGlySerPheLeuGluLeuProGluThrPheCysSerAlase 775
      ::|||::|  |||||:::|||||:::|||||:::|||||
2146 CTAACGGATGCGAGCTTTTGAATCTCTCAGATGAATGTGCCAGGACC 2195
      ::|||::|  |||||:::|||||:::|||||:::|||||
775 rLysProAlaCysGlnGlnAlaCysLysLysAspAspCysProSerGluT 792
      |||:::  :::::|||||  :::  |||||:::|||||:::|||||
2196 CAAGGCATCGTCTCACAAGTCTGTGCCAGGACAGACTGCTCTCCACATT 2245
      ::|||::|  |||||:::|||||:::|||||:::|||||
792 rPLeuLeuSerAspTrpThrGlu 799
      ::|||::|  |||||:::|||||:::|||||:::|||||
2246 TAGCTGTGGGAGATGGTGGAG 2268
      ::|||::|  |||||:::|||||:::|||||:::|||||

```

seq_name: gb_pat:AX235354

seq_documentation_block:

LOCUS AX235354 6294 bp DNA linear PAT 11-SEP-2001

DEFINITION Sequence 1 from Patent WO0162928.

ACCESSION AX235354

VERSION AX235354.1 GI:15593887

KEYWORDS

SOURCE human.

ORGANISM

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 6294)

AUTHORS Vernet,C.A., Fernandes,E., Shimkets,R.A., Macdougall,J. and Spaderna,S.K.

TITLE Polypeptides and nucleic acids encoding same

JOURNAL Patent: WO 0162928-A 1 30-AUG-2001;

Curagen Corporation (US)

FEATURES

source

1..6294

/organism="Homo sapiens"

/db_xref="taxon:9606"

416..4261

/note="unnamed protein product"

/codon_start=1

/protein_id="CAC69811.1"

/db_xref="GI:1559388"

/translation="MPYDFQPLPRWHPNPTACSVSCGGIGRQSRFVCEVBSMHGEI

LQVEWKCWYAPKPKVMQCNLDFCPKWIAMEWSQCTVTCGRGLRYRVVLGINRGEH

VGGNPOILKLIKECVITIPCYKPKESPVKEAKLPWLKQAELEETRIAREETPIIP

EPWASCTTCGQVQVREVKRVLITFTQITELPEECGPKLPTERPCLLEACDES

PASRELDITLPEDSETTYDWEAGFTPCTATCVGQHQAIAVCLHIQQTQVNDSLCD

WVRPPASQACNTEPPEPRHWGVSWCATCGVIGIQTRDVYCLHPGETPAPPEECR

DEKHALQCNQFCPPGWHLIEWQCSRTCCGGTQNRVTCRQLITDGSPLNLSDEL

CKGPKASHKSCATDCPPHILAVGWSKSVSCGVQIRRKQVQORLAAGKRRILPSE

MMCRDLPLCLPVSICOMPCKSIKSEMKTKLGEQGPOTLISVQVYIOTREKRNITI

GSRAYLLPNTSVIILKCPVRRFQKSLIQWEKDGRQLNSKRLGTSGLKIHGLAAPP

IGVYRCIAGSNAQETVVLKLTGTDNRLIARPALREPMREPMGMDHSEANSLSLGVTHHMR

```

OMNKNKNDLYDDDDHISNOPELRALLHCHSNSAGSTNSWELKNKQFEAAVYKQAYSMD
TAQDELIRNMSQLMEGEVSDDLASLIYQIVAEAKAQPTHMQRGIGIQTETPPAAQ
LRGETGSVQSSSHAKNSKLTFKPKGPVLMRQSQPPSTISFNKTIINSRTGNTVYITKRT
EVINILCDLIITSEATYITWTKGTLLOPSVKIILDGPTGKIQTNPTRKEQGIYECSSVA
NHLGSDVRESSVLVYAEAPVILSVERNITKPEHNLSVVGVIYVAAALGANVTIRCPVK
GVQPNITWLKRGSLSGNVSLLFNGLSLLQNVSLNEGTVVCIATNALKXAVATSVL
HLLERRWPESRIVFLOGHKVYLQANTRTNSNDPTGEPPPEPFWEFGNWSHCSATC
GHLGARTORPCQVWANGVESEALCDHLKPLAGFECNIRDCEPARFTSVNSQCSVS
CGEGYHSRQVTCRRTKANGTVQVVSPRACAPKPLGRKPCFGHPYQWEPGNRCPPGR
CMGRAVMQQRHHTACQHNSSDNCDDRRKPTLRNRCTSGACDVCHWHTGPKPKCTAACG
RGFOSRKVDCHITRSCKPVAKRHCQVQKKPISWRHCLGPSRCDRCTDITTHCYMFKVHL
NLCSLDRYKORCCQCOEG*

```

BASE COUNT 1789 a 1297 c 1528 g 1680 t
ORIGIN

alignment_scores:

Quality: 2597.50 Length: 1435

Ratio: 2.851 Gaps: 31

Percent Similarity: 63.484 Percent Identity: 37.561

alignment_block:

US-10-044-807-2 x AX235354 ..

Align seg 1/1 to: AX235354 from: 1 to: 6294

362 SerAspGlyTyrLysGlnIleMetProTyrAspLeuTyrHisProLeuPr 378

|||||:::||||| |||||:::|||||:::|||||:::|||||

395 AGTGATGATTTAAAGAGATAATGCCCTATGACCACCTTCCAACTCTTCC 444

|||||:::||||| |||||:::|||||:::|||||:::|||||

378 oArgTrpGluAlaThrProTyrAlaCysSerSerSerCysGlyGlyG 395

|||||:::||||| |||||:::|||||:::|||||:::|||||

445 TCGCTGGGAACATAATCTTCGACTGCATGTTCCGTCTCTGTGGAGGAG 494

|||||:::||||| |||||:::|||||:::|||||:::|||||

395 lYleGlnSerArgAlaValSerCysValGluGluAspIleGlnGlyHis 411

|||||:::||||| |||||:::|||||:::|||||:::|||||

495 GGATTCAGAGACGGAGCTTTGTGTGTGTAGAGGAATCCATGCATGGAGAG 544

|||||:::||||| |||||:::|||||:::|||||:::|||||

412 ValThrSerValGluGluTrpLysCysMetTyrThrProLysMetProil 428

::: :::::||||| |||||:::|||||:::|||||:::|||||

545 ATATTGCAGTGGGAAGAATGGAAGTGCATGTACGCACCAACCAACCAAGGT 594

|||||:::||||| |||||:::|||||:::|||||:::|||||

428 eAlaGlnProCysAsnIlePheAspCysProLysTrpLeuAlaGlnGluT 445

::: :::::||||| |||||:::|||||:::|||||:::|||||

595 TATGCAAACTTGTAATCTGTTGATGCCCAAGTGGATGCCATGGAGT 644

|||||:::||||| |||||:::|||||:::|||||:::|||||

445 rpSerProCysThrValThrCysGlyGlnGlyLeuArgTyrArgValVal 461

|||||:::||||| |||||:::|||||:::|||||:::|||||

645 GGTCTCAGTGCACAGTACTTGTGGCGAGGAGTTACGGTACCGGGTGTGT 694

|||||:::||||| |||||:::|||||:::|||||:::|||||

462 LeuCysIleAspHisArgGlyMetHisThrGlyGlyCysSerProLysTh 478

|||||:::||||| |||||:::|||||:::|||||:::|||||

695 CTGTGTATTAAACACCCCGGAGAGCATGTTGGGGCTGCAATCCCAACT 744

|||||:::||||| |||||:::|||||:::|||||:::|||||

478 rLysProHisIleLysGluGluCysIleValProThrProCysTyrLysp 495

::: :::::||||| |||||:::|||||:::|||||:::|||||

745 GAAGTTACATCAAAAGAGAATGTGTCATTCCTCCATCCCTGTTTATAAAC 794

|||||:::||||| |||||:::|||||:::|||||:::|||||

495 rLysGluLysLeuProValGluAlaLysLeuProTrpPheLysGlnAla 511

|||||:::||||| |||||:::|||||:::|||||:::|||||

795 CAAAGAAAAAGTCCAGTGGAGCAAAATTCCTTGGCTGGAACAAGCA 844

|||||:::||||| |||||:::|||||:::|||||:::|||||

512 GlnGluLeuGluGluGlyAlaAlaValSerGluGluProSerPheIlePr 528

|||||:::||||| |||||:::|||||:::|||||:::|||||

845 CAGAAGACTAGAAGAGACACAGAAATAGCAACAGAAAGAACCAAGTTCATTC 894

|||||:::||||| |||||:::|||||:::|||||:::|||||

528 oGluAlaTrpSerAlaCysThrValThrCysGlyValGlyThrGlnVala 545

|||||:::||||| |||||:::|||||:::|||||:::|||||

895 AGAACCTTGTCAGCTGACCTACCAAGTGTGGCGCGGTGTGCAGGTCC 944

|||||:::||||| |||||:::|||||:::|||||:::|||||

545 rgtIleValArgCysGlnValLeuSerPheSerGlnSerValAlaAsp 561

::: :::::||||| |||||:::|||||:::|||||:::|||||

945 GTGAGGTGAAGTCCCGTGTCTCTCACATTACGACGACGTGAGCTGAG 994

|||||:::||||| |||||:::|||||:::|||||:::|||||


```

2630 GCTCAGCTCACA.....GGGAAACAGGGAGTGTGCCCAAAG 2667
1159 gProHisArgLys.....ProT 1165
2668 CTCGATGCAAAACACAGCAAGCTGACATTCAGCCGAAAGGACCTG 2717
1165 hrLeuLeuArgLysIleSerAlaAlaGlnGlnLeuSerAlaSerGluVal 1181
2718 TTCATCATGAGCAA...AGCCAACTCCCTCAATTTCATTTAATAAACA 2764
1182 Val...ThrHisLeuGlyGlnThrValAlaLeuAlaSerGlyThrLeuSe 1197
2765 ATAATATCCAGGATGGAAATACAGTATACATTAACAAAAGCACAGAGT 2814
1197 rValLeuLeuHisCysGluAlaIleGlyHisProArgProThrIleSerT 1214
2815 CATCAATATATCTGTGTGACCTTATACCCCGAGTGAGCCACATATACAT 2864
1214 rpAlaArgAsnGlyGluGluValGlnPheSerAspArgIleLeuLeuGln 1230
2865 GGACCAAGGATGAACCTTGTACAGCCCTCAGTAAAAATAATTGGAT 2914
1231 ProAspSerLeuGlnIleLeuAlaProValGluAlaAspValGlyph 1247
2915 GGAACCTGGGAAGATACAGATACAGATCTCTACAAGGAAGAACAGGCAT 2964
1247 eTyrThrCysAsnAlaThrAsnAlaLeuGlyTyrAspSerValSerIleA 1264
2965 ATATGCAATGTCTGTAGCTAATCATCTGTGTCAGATGTGGAAGTCTTT 3014
1264 laValThrLeuAlaGlyLysProLeuValLysThrSerArgMetThrVal 1280
3015 CTGTGCTGTATGACAGACCTGTCTATCTGTGTGTGTAAGAAATATC 3064
1281 IleAsnThrGluLysProAlaValThrValAspIleGlySerThrIleLy 1297
3065 ACCAAACAGACACACCATCTCTCTGTGTGTGTGGAGCATCGTGA 3114
1297 sThrValGlnGlyValAsnValThrIleAsnCysGlnValAlaGlyValp 1314
3115 GGCAGCCCTTGGACCAACAGTGCACATCCGATGTCCTGTAAAGGTGCC 3164
1314 roGluAlaGluValThrTrpPheArgAsnLysSerLysLeu...GlySer 1329
3165 CTCAGCCTAATAATACTTGTGTGAAGAGGAGGAGATCTCTGAGTGGCAAT 3214
1330 ProHisHisLeuHisGluGlySerLeuLeuLeuThrAsnValSerSerSe 1346
3215 GTTTCCTTGTCTTTCATGATGATCCCTGTGTGTGAGAAATGTTCCCTGA 3264
1346 rAspGlnGlyLeuTyrSerCysArgAlaAlaAsnLeuHisGlyGluLeuT 1363
3265 AAATGAAGGAACCTACGTCGTGATGCCACCAATGCTTGTGAAGGCAG 3314
1363 hrGluSerThrGlnLeuLeuIleLeuAspProProGlnValProThrGln 1379
3315 TGGCAACATCTGTACTCCACTTGTGGAAGCAAGATGGCCAGAGAGTAGA 3364
1380 LeuGluAspIleArgAla.....LeuLeuAlaAlaThrGlyPr 1392
3365 ATCGTATTTTCGCAAGGACATAAAAGTACATCTCCAGGCAACCAACAC 3414
1392 oAsnLeuProSerValLeuThrSerProLeuGlyThrGlnLeuValLeuA 1409
3415 TAGAACCAACACAGC.....AATG 3431
1409 spProGlyAsnSerAlaLeuLeuGlyCysProIleLysGlyHisProVal 1425
3432 ACCCA.....ACAGGAGAACCCCG 3451
1426 ProAsnIleThrTrpPheHisGlyGlyGlnProIleValThrAlaThrGI 1442

```

```

3452 CCT..... 3454
1442 yLeuThrHisHisIleLeuAlaAlaGlyGlnIleLeuGlnValAlaAsnL 1459
3454 ..... 3454
1459 euSerGlyGlySerGlnGlyGluPheSerCysLeuAlaGlnAsnGluAla 1475
3454 ..... 3454
1476 GlyValLeuMetGlnLysAlaSerLeuValIleGlnAspTyrTrpPse 1492
3455 .....CAAGAGCCCTTTTGGCA 3471
1492 rValAspArgLeuAlaThrCysSerAlaSerCysGlyAsnArgGlyValG 1509
3472 GCCTGGTAACGTGGTCACATGTGTGCCACCTGTGTTCATTTGGGAGCCC 3521
1509 InGlnProArgLeuArgCysLeuLeu...AsnSerThrGluValAsnPro 1524
3522 GCATTCAGAGACCCACAGTGTGTGATGGCCATGGCGCAGGAAGTGAGTGAG 3571
1525 AlaHisCysAlaGlyLysValArgProAlaValGlnProIleAlaCysAs 1541
3572 GCCCTGTGTGATCACCTCCAGAAGCCACTGGCTGGTGTGAGCCCTGTAA 3621
1541 nArgArgAspCysProSerArgTrpMetValThrSerTrpSerAlaCysT 1558
3622 CATCGGGACTGCCCGAGGTGGTTTCAAAAGTGTGTGGTTCACAGTGT 3671
1558 hrArgSerCysGlyGlyGlyValGlnThrArgArgValThrCysGlnLys 1574
3672 CTGTGCTCTGGGTGAAGATACACACTCGCGAGGTGACGTGCAAGCGG 3721
1575 LeuLysAlaSerGlyIleSerThrProValSerAsnAspMetCysThrGI 1591
3722 ACAAAAGCCAAATGGAACCTGTGCAGGTGTGTCTCCAAGAGCATGTGCC 3771
1591 nValAlaLysArgProValAspThrGlnAlaCysAsnGlnGlnLeuCysv 1608
3772 T...AAAGACCGCCTCTGGGAAGAAACCAATGTTTGTGTCATCATGTG 3818
1608 alGluTrpAlaPheSerSerTrpGlyGlnCysAsnGlyProCysIleGly 1624
3819 TTCAGTGG.....GAACGAGGAACCGGTGTCCTGGAGCTGTCATGGCG 3862
1625 ProHisLeuAlaValGlnHisArgGlnValPheCysGlnThrArgAspGI 1641
3863 CGTGCTGTGAGGATGCAGCGCTCACACAGCTTGTCAACACACAGCTC 3912
1641 yIleThrLeuProSerGluGlnCysSerAlaLeuProArgProValSerT 1658
3913 T.....GACTCCACTGTGATGACAGAAAGAGACCCACCTTAA 3950
1658 hrGlnAsnCysTrpSerGluAlaCysSerValHisTrpArgValSerLeu 1674
3951 GAAGAACTGCACATCAGGCGCTGTGTATGTGTGTGGCACACAGGCCCT 4000
1675 TrpThrLeuCysThrAlaThrCysGlyAsnTyrGlyPheGlnSerArgAr 1691
4001 TGGAGGCCCTGTACAGCAGCTGTGGCAGG...GOTTTCCAGTCTCGAA 4047
1691 qValGluCysValHisAlaArgThrAsnLysAlaValProGluHisLeuc 1708
4048 ACTGACTGTATCCACACAGAGAGTTCAAACCTGTGGCCCAAGAGACACT 4097
1708 ysSerTrpGlyProArgProAlaAsnTrpGlnArgCysAsnIleThrPro 1724
4098 GTGTACAGAAAAAGAAACCAATTTCCTGGCGCACTGT...CTTGGGCC 4144
1725 CysGluAsnMetGluCysArgAspThrThrArgTyrCysGluLysVally 1741
4145 TCCTGTGATAGACTGCACAGACACAACTCACTACTGTATGTGTGTAA 4194

```



```

1606 AGACCAAGGCATGCTCACAAGTCTGTGCCAGGACACTGTCTCTC 1655
790 e rGluTrpLeuSerAspTrpThrGluCysSerThrSerCysGlyGlu 806
1656 CACATTAGCTGGGAGACATGGTCGAAGTGTCAGTGTGGGT 1705
807 GlyThrGlnThrArgSerAlaIleCysArgLysMetLeuLysThrGlyLe 823
1706 GGAATCCAGAGAAGAAAGCAGGTGTGCAAAAGGCTGCAGCCAAAGTCG 1755
823 uSerThrValValAsnSerThrLeuCysProProLeuProPheSerSers 840
1756 GCGCATCCCTCAGTCAGATGATGTCAGGGATCTACAGGGTCCCTC 1805
840 erLeArgProCysMetLeuAlaThrCysAlaArg...ProGlyArgPro 855
1806 TTGTAAGATCTGCCAGATGCCCTCAGTCAGTAAATCAATCAGAGATG 1855
856 SerThrLys.....HisSerProHisIleAlaAlaAlaArgLysVa 869
1856 AAGACAAAACCTTGTGTAGCAGGGTCCGAGATCCTCAGTGTCCAGAGAT 1905
869 lTyrIleGlnThrArgArgGlnArgLysLeuHisPheValValGlyGlyP 886
1906 CTACATTCAGACAGGAAGAGAGAGCGTATTAACTGACCATGGTAGCA 1955
886 heAlaTyrLeuLeuProLysThrAlaValValLeuArgCysProAlaArg 902
1956 GAGCTATTGCTGCCCAACACATCCGTGATTATTAGTCCCGTGGCA 2005
903 ArgValArgLysProLeuIleThrTrpGluLysAspGlyGlnHisLeuIl 919
2006 CGATTCCAGAAATCTCTGATCCAGTGGGAGAGGATGCCGCTTGCCTGCA 2055
919 eSerSerThrHisValThrValAlaProPheGlyTyrLeuLysIleHisA 936
2056 GAATCCAAACGGTGGCATGCCAAGTCCAGGCTCACTAAATCCACG 2105
936 rgLeuLysProSerAlaGlyValTyrThrCysSerAlaGlyProAla 952
2106 GTCTGTGCTCCCGCATCGCGGTGTACCGGTGCATGTCAGGCTTCGCA 2155
953 ArgGluHisPheValIleLysLeuIleGlyClyAsnArgLysLeuValAl 969
2156 CAGGAACAGTTGTGCTCAAGCTCATTTGGTACTGACACCGGCTCATCGC 2205
969 aArgProLeuSerProArgSerGluGluGluValLeuAlaGlyArgLysG 986
2206 ACGCCCA.....GCCCTCAGGGAGC 2225
986 lyGlyProLysGluAlaLeuGlnThrHisLysHisGlnAsnGlyIlePhe 1002
2226 CTATGAGGAATATCCGGCATGCCAGCACAGCGCAATAGTTTGGGA 2275
1003 SerAsnGlySerLysAlaGluLysArgGlyLeuAlaAlaAsnProGlySe 1019
2276 GTCACATGGCACAAA.....ATGAGGCAAAATGTGGAATAACAAAATGA 2319
1019 rArgTyr.....AspAspLeuValSerArg.....L 1028
2320 CCTTATCTCGATGAGCACCACATAGTAAACCAGCCCTTCTTGAGAGCTC 2369
1028 euLeuGluGlnGlyIleTrpProGlyGluLeuAlaSerTrpGluAla 1044
2370 TGTTAGGCCACTGCAGCAATCTGCAGGAGCACCAACTCTCGGAGTTG 2419
1045 GlnAspSerAlaGluArgAsnThrThrSerGluGluAspProGlyAlaGl 1061
2420 AAGAAATAGCAGTTTGAAGCAGCAGGTTAAACAA.....GGAGCA.. 2458
1061 uGlnValLeuLeuHisLeuProPheThrMetValThrGluGlnArgL 1078

```

```

2459 .....TATAGCATGGATACAGCCAG.....T 2480
1078 euAspAspIleLeuGlyAsnLeuSerGlnGlnProGlu.....GluLeu 1092
2481 TTGAIGAGCTGATAGAAACATGAGTCAGCTCATGGAAACCGGAGAGTCT 2530
1093 ArgAspLeuTyrSerLysHisLeuValAlaGlnLeuAlaGlnGluIlePh 1109
2531 AGCGATGATCTTGGTCCAGCTGATATATCAGTGTGGCCGATATTAGC 2580
1109 eArgSerHisLeuGluHisGlnAspThrLeuLeuLysProSerGluArgA 1126
2581 CAAGGCACAGCCAAACACACATGTCAG.....TGGCGGGCATCCAGGAAG 2624
1126 rgThrSerProValThrLeuSerProHisLysHisValSerGlyPheSer 1142
2625 AGACACCTCCT.....GCT 2638
1143 SerSerLeuArgThrSerSerThrGlyAspAlaGlyGlyGlySerArgAr 1159
2639 GCTCAGCTCAGA.....GGGAAACAGGAGTGTGTCCCAAG 2676
1159 gProHisArgLys.....ProT 1165
2677 CTCGCATGCAAAAAACTCAGCAAGCTGACATTCAAGCCGAAAGGACCTG 2726
1165 hrIleLeuArgLysIleSerAlaAlaGlnGlnLeuSerAlaSerGluVal 1181
2727 TTCATGAGGCAA...AGCCAACCTCCCTCANTTTCATTTAATAAACA 2773
1182 Val...ThrHisLeuGlyGlnThrValAlaLeuAlaSerGlyThrLeuSe 1197
2774 ATAATATCCAGGATTGGAAATACAGTATACATTACAAAAGACAGCAGGT 2823
1197 rValLeuLeuHisCysGluAlaIleGlyHisProArgProThrIleSerT 1214
2824 CATCAATATACTGTGTGACCTTATTACCCCGAGTGAGGCCACATACAT 2873
1214 rpAlaArgAsnGlyGluGluValGlnPheSerAspArgIleLeuLeuGln 1230
2874 GGACCAAGGATGGAACTTGTTCAGCCCTCAGTAAATAATTTTGGAT 2923
1231 ProAspAspSerLeuGlnIleLeuAlaProValGluAlaAspValGlyPh 1247
2924 GGAATCGGAAGATACAGATACAGATCCTACAGGAAGAAACAGGCAT 2973
1247 eTyrThrCysAsnAlaThrAsnAlaLeuGlyTyrAspSerValSerIleA 1264
2974 ATATGAATGTCTGTAGCTAATCATCTTGTTCACATGTGGAAGTTCTT 3023
1264 laValThrLeuAlaGlyLysProLeuValLysThrSerArgMetThrVal 1280
3024 CTGTGCTGTATGAGAGCCCTGTCACTGTCTGTGTGAAGAAATATC 3073
1281 IleAsnThrGluLysProAlaValThrValAspIleGlySerThrIleLy 1297
3074 ACCAAACAGAGACACCACTCTCTGTGTGTGTGGAGCATCGTGA 3123
1297 sThrValGlnGlyValAsnValThrIleAsnCysGlnValAlaGlyValP 1314
3124 GGCAGCCCTTGGAGCAACCGTGACAATCCGATGCTCTGTAAGAGGTGTC 3173
1314 roGluAlaGluValThrTrpPheArgAsnLysSerLysLeu...GlySer 1329
3174 CTCAGCCTAATAATACTTGGTTGAAGAGAGGAGATCTCTGAGTGGCAAT 3223
1330 ProHisHisLeuHisGluGlySerLeuLeuLeuThrAsnValSerSe 1346
3224 GTTTCCTTGTCTTTCATAGGATCCCTGTGTGTGCAGAATGTTTCCCTGA 3273
1346 rAspGlnGlyLeuTyrSerCysArgAlaAlaAsnLeuHisGlyGluLeuT 1363
3274 AAATGAAGGAACCTTACGTCTGTAGTACCAATGCTCTTGGAAAGGCAG 3323

```

```

1363 hrGluSerThrGlnLeuLeuLeuLeuAspProGlnValProThrGln 1379
  :: ::::: ||| ::::: ||| ::::: ||| ::::: ||| ::::: |||
3324 TGGCAACATCTGTACTCTCACTTGTGGAAGAAAGATGCCAGAGATAGA 3373
1380 LeuCluaspIleArgAla.....LeuLeuAlaAlaThrGlyPr 1392
  :: ::::: ||| ::::: ||| ::::: ||| ::::: ||| ::::: |||
3374 ATCGTATTTCTGAAGGACATAAAAGTACATTCCTCCAGCAACCAACAC 3423
1392 oAsnLeuProSerValLeuThrSerProLeuGlyThrGlnLeuValLeuA 1409
  :: ::::: ||| ::::: ||| ::::: ||| ::::: ||| ::::: |||
3424 TAGAACCAACAGC.....AATG 3440
1409 spProGlyAsnSerAlaLeuLeuGlyCysProIleLysGlyHisProVal 1425
  |||||
3441 ACCCA.....ACAGGAGAACCCCG 3460
1426 ProAsnIleThrPheHisGlyGlnProIleValThrAlaThrG1 1442
  |||
3461 CCT..... 3463
1442 yLeuThrHisIleLeuAlaAlaGlyGlnIleLeuGlnValAlaAsnL 1459
3463 ..... 3463
1459 euSerGlyGlySerGlnGlyCluPheSerCysLeuAlaGlnAsnGluAla 1475
3463 ..... 3463
1476 GlyValLeuMetGlnLysAlaSerLeuValIleGlnAspTyrTrpTrpSe 1492
3464 .....CAAGAGCCTTTTGGGA 3480
1492 rValAspArgLeuAlaThrCysSerAlaSerCysGlyAsnArgGlyValG 1509
  :: ::::: ||| ::::: ||| ::::: ||| ::::: ||| ::::: |||
3481 GCCTGGTAACGTGTCACATGTTCTGCCACCTGTGGTCATTTGGGAGCCC 3530
1509 lnGlnProArgLeuArgCysLeu.....AsnSerThrGluValAsnPro 1524
  :: ::::: ||| ::::: ||| ::::: ||| ::::: ||| ::::: |||
3531 GCAFTTCAGAACCCCGAGTGTGTATGTCCTCAATGGCAGGAAGTGAGTGAG 3580
1525 AlaHisCysAlaGlyLysValArgProAlaValGlnProIleAlaCysAs 1541
  ||| |||
3581 GCCCTGTGTATACCTCCAGAGAGCCACTGGCTGGGTGTGAGCCCTGTAA 3630
1541 nArgArgAspCysProSerArgTrpMetValThrSerTrpSerAlaCyst 1558
  |||||
3631 CATCCGGCACTGCCAGCGAGGTGTTCAACAAGTCTGTGTCACAGTGCT 3680
1558 hrArgSerCysGlyGlyValGlnThrArgArgValThrCysGlnLys 1574
  :: ||||| ||| ::::: ||| ::::: ||| ::::: ||| ::::: |||
3681 CTGTGCTTTCGGTGAAGGATACCACAGTCGGCAGGTGACGTGCAGCGG 3730
1575 LeuLysAlaSerGlyIleSerThrProValSerAsnAspMetCysThrG1 1591
  |||||
3731 ACAAAACCAATGGAACGTGTGAGGTGCTCTCCAGAGCATGTGCCCC 3780
1591 nValAlaLysArgProValAspThrGlnAlaCysAsnGlnGlnLeuCysV 1608
3781 T...AAAGACCGGCTCTGGGAAGAAAACCATGTTTGGTCATCATGTG 3827
1608 alGluTPAlaPheSerSerTrpGlyGlnCysAsnGlyProCysIleGly 1624
  |||||
3828 TTCAGTGG.....GAACAGGAAACCGGTGCTCTGGACGTGTGATGGG 3871
1625 ProHisLeuAlaValGlnHisArgGlnValPheCysGlnThrArgAspG1 1641
  :: ::::: ||| ::::: ||| ::::: ||| ::::: ||| ::::: |||
3872 CGTCTGTGAGGATGCAGAGCGTCACACAGCTTGTCAACACACAGCTC 3921
1641 yIleThrLeuProSerGlnCysSerAlaLeuProArgProValSerT 1658
  :: ::::: ||| ::::: ||| ::::: ||| ::::: ||| ::::: |||
3922 T.....CACTCCAACTGTGATGACAGAAAGAGACCCACTTAA 3959

```

```

1658 hrGlnAsnCysTrpSerGluAlaCysSerValHisTrpArgValSerLeu 1674
  :: ::::: ||| ||||| ::::: ||| ::::: ||| ::::: |||
3960 GAAGGATCGCATCATGAGGGCGCTGTATGTGTGTGGCACACAGGCCT 4009
1675 TrpThrLeuCysThrAlaThrCysGlyAsnTyrGlyPheGlnSerArgAr 1691
  ||| ||||| ::::: ||| ::::: ||| ::::: ||| ::::: |||
4010 TGAAGCCCTGTACAGCAGCCTGTGGCAGG...GCTTCCAGTCTCGGAA 4056
1691 gValGluCysValHisAlaArgThrAsnLysAlaValProGluHisLeuC 1708
  :: ::::: ||| ::::: ||| ::::: ||| ::::: ||| ::::: |||
4057 AGTCACTGTATCCACACACAGAGGTGCAAAACCTGTGGCCCAAGAGACT 4106
1708 ysSerTrpGlyProArgProAlaAsnTrpGlnArgCysAsnIleThrPro 1724
  || ::||| ::||| ::||| ::||| ::||| ::||| ::|||
4107 GTGTACAGAAAAGAAACCAATTCCTGGCGGCACTGT...CTTGGGCC 4153
1725 CysGluAsnMetGluCysArgAspThrArgTyrCysGluLysValLy 1741
  :: ::::: ||| ||||| ::||| ::||| ::||| ::|||
4154 TCCTGTGATAGAGACTGCACAGACACAACCTCACTACTGTATGTGTAAA 4203
1741 sGlnLeuLysLeuCysGlnLeuSerGlnPheLysSerArgCysCysGlyT 1758
  ||||| ::||| ::||| ::||| ::||| ::||| ::||| ::|||
4204 ACATCTTAATTGTGTTCTCTAGACCGCTACAAACAAGGTGCTGCCAGT 4253
1758 hrCys 1759
  ::|||
4254 CATGT 4258

seq_name: gb_pr:AF251058
seq_documentation_block:
LOCUS AF251058 1616 bp mRNA linear PRI 15-APR-2001
DEFINITION Homo sapiens clone 2 thrombospondin mRNA, complete cds.
ACCESSION AF251058
VERSION AF251058.1 GI:13625177
KEYWORDS
SOURCE
  ORGANISM
    Homo sapiens
    Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
    Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
  1 (bases 1 to 1616)
AUTHORS
  Mao,Y., Xie,Y., Zhou,Z., Zhao,W., Zhao,S., Wang,W., Huang,Y.,
  Wang,S., Tang,R., Chen,X. and Wu,C.
TITLE
  Direct Submission
JOURNAL
  Submitted (29-MAR-2000) Institute of Genetics, School of Life
  Sciences, Fudan University, 220 Handan Road, Shanghai 200433, P.R.
  China
FEATURES
  source
    1..1616
      /organism="Homo sapiens"
      /db_xref="taxon:9606"
      /clone="2"
      162..1415
        /codon_start=1
        /product="thrombospondin"
        /protein_id="AAK34948.1"
        /db_xref="GI:13625178"
  CDS
    162..1415
      /translation="MAGPLTADFIVKIRNSGADSTVQFIYQPIIHRWRETDFFPCS
      ATCGGYQLTSAECYDLRSNRVWQYCHYIPENIKPKFKLOECNLDPCPASDGYKQI
      MPYDLYPLPRWFATPTWATSSCGGIGQISRVCSVEEDIOGHVTSVEWKMCMYTKM
      PIAQCNIFDCPKWLAQWSPCTVTCGGIYRVRVLCIDHRGMHTGCCSKPKPHIKE
      ECIVTPCYKPKELPFAKLPRFKQAQELBEGAAYSESPFIPEAWSACTVTCGVGT
      QVRIVRCVLLSFSQVADLPIDECGPKPACAGTACGSCGETPEFNPDETDGLFG
      GLODFEYDWEYETGKCSGCGGVGAIVVSLNKTQPREAENLCVTSRRPPOLL
      KSCNLDPCPARSIDSAWNAACNVLC"
BASE COUNT 386 a 438 c 426 g 366 t
ORIGIN

```

```

alignment_scores:
  Quality: 2558.00      Length: 534
  Ratio: 5.242         Gaps: 3
  Percent similarity: 91.386  Percent identity: 87.828

```


alignment_block:

US-10-044-807-2 x AF251058 ..

Align seg 1/1 to: AF251058 from: 1 to: 1616

```
214 HisIleArgLeuValLeuLysGlyProAspHisLeuTyrLeuGluThrLy 230
|||||
3 CATATTCGCTTGTCTTAAAGGCTCTGATCATTATATCTGGAACCAA 52
230 sThrLeuGlnGlyThrLysGlyGluAsnSerLeuSerSerThrGlyThrP 247
|||||
53 AACCTCCAGGGGACTAAAGGTGAACAGTCTCAGCTCCACAGGAAGTT 102
247 heLeuValAspAsnSerSerValAspPheGlnLysPheProAspLysGlu 263
|||||
103 TCCITGTGACAAATTCATGTGGACTTCCAGAAATTTCCAGACAAAGAG 152
264 IleLeuArgMetAlaGlyProLeuThrAlaAspPheIleValLysIleAr 280
|||||
153 ATACTGAGAATGGCTGGACCACTCACAGCAGATTTCATGTCAAGATTGC 202
280 gAsnSerGlySerAlaAspSerThrValGlnPheIlePheTyrGlnProI 297
|||||
203 TAATCTGGGCTCCGCTGACATGACAGTCCAGTTCATCTTATCAACCCA 252
297 leIleHisArgTyrArgGluThrAspPhePheProCysSerAlaThrCys 313
|||||
253 TCATCCACCGATGGAGGAGACGGATTTCCTTCTGCTCAGCAACCTGT 302
314 GlyGlyGlyTyrGlnLeuThrSerAlaGluCysTyrAspLeuArgSerAs 330
|||||
303 GGAGGAGGTTATACAGCTGACATCGCTGAGTGCTACGATCTGAGGAGCAA 352
330 nArgValValAlaAspGlnTyrCysHisTyrTyrProGluAsnIleLysP 347
|||||
353 CCGTGTGTTCTGCTGACCAATACTGTCACTATTACCCAGAGAACATCAAC 402
347 roLysProLysLeuGlnGluCysAsnLeuAspProCysProAlaSerAsp 363
|||||
403 CCAAAACCAAGCTTCAGGAGTGCAACTTGGATCTCTTGCCAGGCCAGTCA 452
364 GlyTyrLysGlnIleMetProTyrAspLeuTyrHisProLeuProArgTyr 380
|||||
453 GGATACAGCAGATCATGCCCTTATGACCTCTACCATCCCTTCCTCGGTG 502
380 pGluAlaThrProTyrThrAlaCysSerSerCysGlyGlyGlyIleG 397
|||||
503 GGAGGCCACCCCATGGACCGGTGCTCTCTCTGCTGGGGGGGGCATCC 552
397 InSerArgAlaValSerCysValGluGluAspIleGlnGlyHisValThr 413
|||||
553 AGAGCCGGGCAGTTTCCTGTGTGGAGGAGGACATCCAGGGGCGCTCACT 602
414 SerValGluGluTyrLysCysMetTyrThrProLysMetProIleAlaGl 430
|||||
603 TCAGTGAAGAGTGAATGATGATACACCCCTAAGATGCCATCGCCATCGCCA 652
430 nProCysAsnIlePheAspCysProLysTyrLeuAlaGlnGluTyrSerP 447
|||||
653 GCCCTGAACAATTTTGACTGCCCTNAATGGCTGGCAGAGAGTGGTCTC 702
447 roCysThrValThrCysGlyGlnGlyLeuArgTyrArgValValLeuCys 463
|||||
703 CGTGACAGTGACATGTGGCAGGGGCTTCAGATACCGGTGCTCTCTCTGC 752
464 IleAspHisArgGlyMetHisThrGlyCysSerProLysThrLysPr 480
|||||
753 ATCGACCATCAGGAATGCACACAGGAGGCTGTAGCCCAAAACAAAGGCC 802
480 oHisIleLysGluGluCysIleValProThrProCysTyrLysProLysG 497
|||||
803 CCACATAAAGAGGATGCATCGTACCCACTCCCTGCTATATAAACCCAAAG 852
```

```
497 luLysLeuProValGluAlaLysLeuProTyrPheLysGlnAlaGlnGlu 513
|||||
853 AGAAACTTCCAGTCAGGCCCAAGTTGCCATGGTTCAACAACAGCTCAAGAG 902
514 leuGluGluGlyAlaAlaValSerGluGluProSerPheIleProGluAl 530
|||||
903 CTAGAAAGAGGAGCTGTGTGTGACAGGAGCCCTCGTTTCATCCAGAGGC 952
530 aTyrPheAlaCysThrValThrCysGlyValGlyThrGlnValArgIleV 547
|||||
953 CTGGTGGCCCTGCACAGTCACTGTGTGTGGGACCCAGGTGCGAATAG 1002
547 alArgCysGlnValLeuLeuSerPheSerGlnSerValAlaAspLeuPro 563
|||||
1003 TCAGTGGCAGGTGCTCTGTCTTCTCTCAGTCCGTGGCTGACCTGCCT 1052
564 IleAspGluCysGluGlyProLysProAlaSerGlnArgAlaCysTyrAl 580
|||||
1053 ATTGACGAGTGTGAAGGGCCCAAGCCAGCATCCAGCGTGCCTGTTATGC 1102
580 acLysProCysSerGlyGluIleProGluPheAsnProAspGluThrAspG 597
|||||
1103 AGGCCCATGCAGCGGGGAAATTCCTGAGTCAACCCAGACGAGACAGATG 1152
597 lyLeuPheGlyGlyLeuGlnAspPheAspGluLeuTyrAspTyrGluTyr 613
|||||
1153 GGCTCTTGGTGGCCTGCGAGGATTTCCGAGGCTGTATGACTGGAGATAT 1202
614 GluGlyPheThrLysCysSerGluSerCysGlyGlyValGlnGluAl 630
|||||
1203 GAGGGGTTTACCAAGTGTCCGAGTCTGTGGAGGAGGTGTCCAGAGGC 1252
630 aValValSerCysLeuAsnLysGlnThrArgGluProAlaGluGluAsnL 647
|||||
1253 TGTGTGAGCTGCTTGAACAAACACACTCGGAGGCTGCTGAGGAGAACCC 1302
647 euCysValThrSerArgArgProGlnLeuLeuLysSerCysAsnLeu 663
|||||
1303 TGTGGGTGACGAGCGCGCGGCCCCACAGCTCTCTGAAGTCTTGCAATTG 1352
664 AspProCysProAlaArgTyrGluIle...GlyLysTyrSerProCysSe 679
|||||
1353 GATCCCTGCCAGCAAGAAGAGTATCGACTCAGCATGGAGCGCTGCAA 1402
679 rLeuThrCysGlyValGlyLeuGlnThrArgAspValPheCysSerHisL 696
|||||
1403 CTTCTTTGTAGGCAACCAAGAGGCGCTGCTCTCATCTCTGCTGTCCACC 1452
696 euLeuSer.....Arg 699
|||||
1453 AACTAGCTCTGGGCTAGGGCGAGGTGTCTGCCCTTATGTTTCCACAT 1502
700 GluMetAsnGluThrValIleLeuAlaAspGluLeuCysArgGlnProLy 716
|||||
1503 CTGCAAGTGAACCTGGTTGTA..... 1523
716 sProSerThrValGlnAlaCysAsnArgPheAsnCysProProAlaTyrP 733
|||||
1524 .CCTGATGATCGAGATCCCATGACTTGTCTCATATGCCATGATTCCTT 1572
733 yr 733
1573 AT 1574
```

seq_name: gb_pat:AX224815

seq_documentation_block:

LOCUS AX224815 2175 bp DNA linear PAT 10-SEP-2001
DEFINITION Sequence 7 from Patent WO0161011.
ACCESSION AX224815
VERSION AX224815.1 GI:15554912
KEYWORDS


```
1640 TGGCTTGGCTGAAACAAGCAAGCAAGTACAGAGACCAAGTACAGCAAC 1689
|||||
522 GluGluProSerPheIleProGluAlaIlePheSerAlaCysThrValThrCy 538
|||||
1690 GAAGAACCACAGTTCATTCAGAACACCTGGTACAGCTGCAGTACCAAGCTG 1739
|||||
538 sGlyValGlyThrGlnValArgIleValArgCysGlnValIleLeuSerP 555
|||||
1740 TGGCCAGGTGTGTCAGGTCCGCGAGGTGAAGTCCGCTGTCTCTCAT 1789
|||||
555 heSerGlnSerValAlaAspLeuProIleAspGluCysGluGlyProIlys 571
|||||
1790 TCAGGCAGACTGAGACTGAGCTGCCGAGGAGAGTGTGAAGCCCAAG 1839
|||||
572 ProAlaSerGlnArg.....AlaCystyrAlaGlyProCy 583
|||||
1840 CTGCCACCAAGCGCCCTGCCCTCGGAGCATGTGATGAGAGCCCGC 1889
|||||
583 sSerGlyGluIleProGluPheAsnProAspGluThrAspGlyLeuPheG 600
|||||
1890 CTCGAGAGCTAGACTCCCTCTCCCT..... 1917
|||||
600 LyGlyLeuGlnAspPheAspGluLeuTyrrAspTrpGluTyrrGluGlyPhe 616
|||||
1918 .....GAGGACAGTACAGACGACTTACGACTGGGAGTACGCTGGGTT 1959
|||||
617 ThrLysCysSerGluSerCysGlyGlyValGlnGluAlaValValse 633
|||||
1960 ACCCTTGCACAGCAACATCTTGGGAGGCCATCAAGAGCCATAGCAGT 2009
|||||
633 rCysLeuAsnLysGlnThrArgGluProAlaGluGluAsnLeuCysValt 650
|||||
2010 GTGCTTACATATCAGACCCAGCAGCAGTCAATGACAGCTTGTGTGATA 2059
|||||
650 hrSerArgArgProGluLeuLysSerCysAsnLeuAspProCys 666
|||||
2060 TGTCCACCTCTCCAGCATGAGCCAGCGCTGTAACACAGAGCCCTGT 2109
|||||
667 ProAlaArgTrpGlu 671
|||||
2110 CCCCCCAGGAGAGAG 2124
|||||
```

seq_name: gb_pat:AX149471

seq_documentation_block: 2912 bp DNA linear PAT 08-JUN-2001
LOCUS AX149471
DEFINITION Sequence 27 from Patent W00136638.

ACCESSION AX149471

VERSION AX149471.1 GI:14347941

KEYWORDS

SOURCE human.

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1. (bases 1 to 2912)

AUTHORS Shimkets, R.A., Lichenstein, H., Vernet, C. and Fernandes, E.

TITLE Polypeptides and nucleic acids encoding same

JOURNAL Patent: WO 0136638-A 27 25-MAY-2001;

Curagen Corporation (US)

FEATURES

Location/Qualifiers

1..2912

/organism="Homo sapiens"

/db_xref="taxon:9606"

BASE COUNT 810 a 713 c 754 g 635 t

ORIGIN

alignment_scores:

Quality: 2346.50 Length: 1023

Ratio: 3.273 Gaps: 21

Percent Similarity: 70.088 Percent Identity: 44.184

alignment_block:

US-10-044-807-2 x AX149471

Align seg 1/1 to: AX149471 from: 1 to: 2912

```
240 SerLeuSerSerThrGlyThrPheLeuValAspAsnSerSerValAspPh 256
|||||
2 AGCTTTAACAGCCCGCGCTTTGTCGTAGAAAAACACACAGTGAATT 51
|||||
256 eGlyLysPheProAspLysGluIleLeuArgMetAlaGlyProLeuThrA 273
|||||
52 TTAGAGGGGCTCCGAGAGGCAAACTTTTAAAGATTCCAGGCCCTTTGATGG 101
|||||
273 laAspPheIleValLysIleArgAsnSerGlySerAlaAspSerThrVal 289
|||||
102 CTGATTTCATCTTCAAGACAGGTACACTGCAGCCAAAGACAGCGGTGTT 151
|||||
290 GlnPheIlePheThrGlnProIleIleHisArgTrpArgGluThrAspPh 306
|||||
152 CAGTTCTTCTTTTACCAGCCCATCAGTCATCAGTGGGACACAACTGACTT 201
|||||
306 ePheProCysSerAlaThrCysGlyGlyThrGlnLeuThrSerAlaG 323
|||||
202 CTTTCCCTGTCACCTGTGAGCTGTGGAGAGGTATCAGCTCAATTCTGCTG 251
|||||
323 luCysTyrrAspLeuArgSerAsnArgValValAlaAspGlnTyrrCysHis 339
|||||
252 AATGTGTGGATATCCGCTTGAAGAGGTAGTTCCTGACCATATATGTCAC 301
|||||
340 TyrTyrrProGluAsnIleLysProLysLeuGlnGluCysAsnLe 356
|||||
302 TACVACCTGAAATGTAAAAACCAAACTGAAAGTAAAGGAAATGCAGCAT 351
|||||
356 uAspProCysProAlaSerAspGlyTyrrLysGlnIleMetProTyrrAspL 373
|||||
352 GGATCCCTGCCATCAGTGTGATGATTTAAAGAGATTAATGCCCTATGACC 401
|||||
373 euTyrrHisProLeuProArgTrpGluAlaThrProTrpThrAlaCysSer 389
|||||
402 ACTTCCAACTCTTCCCTCGCTGGGAACATATCTTGGACTGCATGTCC 451
|||||
390 SerSerCysGlyGlyIleGlnSerArgAlaValSerCysValGluG 406
|||||
452 GTGTCCTGTGGAGGAGGATTCAGAGACGGAGCTTTGTGTGTGTAGAGGA 501
|||||
406 uAspIleGlnGlyHisValThrSerValGluGluTrpLysCysMetTyrr 423
|||||
502 ATCCATGCATGGAGAGATATTCAGGTGGAAGATGGAAGTCCATGTACG 551
|||||
423 hrProLysMetProIleAlaGlnProCysAsnIlePheAspCysProLys 439
|||||
552 CACCCAAACCCCAAGTTATGCAAACTTGTATCTGTTGATGCCCCAAG 601
|||||
440 TrpLeuAlaGlnGluTrpSerProCysThrValThrCysGlyGlnGlyLe 456
|||||
602 TGGATTGCCATGGAGTGTCTCAGTCAGCAGTGCAGTGTGTCGCCAGGTT 651
|||||
456 uArgTyrrArgValValLeuCysIleAspHisArgGlyMetHisThrGlyG 473
|||||
652 ACGGTACCGGGTGTCTGTGTATTAACCCACCGCGGAGAGCATGTTGGGG 701
|||||
473 lyCysSerProLysThrLysProHisIleLysGluGluCysIleValPro 489
|||||
702 GCTGCAATCCACCACTGAGTTACATCAAGAGAGATGTGTCATTTCCC 751
|||||
490 ThrProCysTyrrLysProLysGluLysLeuProValGluAlaLysLeuPr 506
|||||
752 ATCCCGTGTATTATAAACCAAAAGAAAAAGTCCAGTGAAGCAAAATTTGCC 801
|||||
506 oTrpPheLysGlnAlaGlnGluLeuGluGluGlyAlaAlaValSerGluG 523
|||||
802 TTGGCTGAAACAGCAAGAACTAGAAAGACCAAGCAAGTAAAGCAACAGAA 851
|||||
```

```

523 luproSerPheIleProGluAlaTrpSerAlaCysThrValThrCysGly 539
852 AACCAAGTTTCATCCAGAACCTGGTCAGCTGCAGTACCACTGTGGG 901
540 ValGlyThrGlnValArgIleValArgCysGlnValLeuLeuSerPheSe 556
902 CCGGTGTGCAGGTCCGTGAGTGAAGTGCCTGCTCTCATATTAC 951
556 rGlnSerValAlaAspLeuProIleAspGluCysGluGlyProLysProA 573
952 GCAGACTGAGACTGAGTCCCGGAGGAAGAGTGTGAAGGCCCAAGCTGC 1001
573 laSerGlnArg.....AlaCysTyrAlaGlyProCysSer 584
1002 CCACCGAACCGCCCTGCTCTCTGGAAGCATGTGATGAGAGCCCGGCTCC 1051
585 GlyGluIleProGluPheAsnProAspGluThrAspGlyLeuPheGlyG 601
1052 CGAGAGCTACATCCCTCTCCT..... 1075
601 yLeuGlnAspPheAspGluLeuTyrAspTrpGluTyrGluGlyPheThrL 618
1076 ....GAGGACAGTGCAGACGACTTACGAGCTGGAGTACGCTGGGTTCACC 1121
618 yscysSerGluSerCysGlyGlyValGlnGluAlaValSerCys 634
1122 CTTGCACAGCAACATGCTGGGAGGCCATCAAGAAGCCATGAGTAGTGC 1171
635 LeuAsnLysGlnThrArgGluProAlaGluGluAsnLeuCysValThrSe 651
1172 TTACATATCCAGACCCAGACAGTCAATGACAGCTGTGTGATATGGT 1221
651 rArgArgProGlnLeuLeuLysSerCysAsnLeuAspProCysProA 668
1222 CCACCGTCTCCAGCCATGAGCAGGCGCTTAACACAGAGCCCTGTCCCC 1271
668 laArgTrpGluIleGlyLysTrpSerProCysSerLeuThrCysGlyVal 684
1272 CCAGTGGCATGTGGGCTCTGGGGCCCTGCTCAGTACCTGTGGAGTT 1321
685 GlyLeuGlnThrArgAspValPheCysSerHisLeuLeuSerArgGluMe 701
1322 GGAATTCAGACCCGAGATGTACTGCTGCAC.....CC 1356
701 tAsnGluThrValIleLeuAlaAspGluLeuCysArgGlnProLysProS 718
1357 AGGGAGAGCCCTGCCCTCTCTGAGGAG...TGCCGAGATGAAGAAGCCCC 1403
718 erThrValGlnAlaCysAsnArgPheAsnCysProProAlaTtpTyrPro 734
1404 ATGCTTTACAGCAGTCGATCAGTTTACCTGCTCTCTGCTGGCATT 1453
735 AlaGlnTrpGlnProCysSerArgThrCysGlyGlyValGlnLysAr 751
1454 GAAGATGGCAGCAGTGTTCAGGACTGTGGGGGGAACTCAGAACAG 1503
751 gGluValLeuCysLysGlnArgMetAlaAspGlySerPheLeuLeuLeup 768
1504 AAGAGTCACCTGTGGCAGCTGTAAAGCGATGGCAGCTTTTGAATCTCT 1553
768 roGluThrPheCysSerAlaSerLysProAlaCysGlnAlaCysLys 784
1554 CAGATGAATGTGGCAGGACCCAGGCATCGTCTCAAGTCTGTGCC 1603
785 LysAspAspCysProSerGluTrpLeuLeuSerAspTrpThrGluCysSe 801
1604 AGGACAGACTCTCTCCACATTTAGCTGTGGGAGCTGGTGAAGTCTC 1653
801 rThrSerCysGlyGluGlyThrGlnThrArgSerAlaIleCysArgLysM 818
1654 TGTCAAGTTGTGGTGTGGAAATCCAGAGAAAGACAGGTGTGTCAAAAGC 1703
818 etLeuLysThrGlyLeuSerThrValValAsnSerThrLeuCysProPro 834

```

```

1704 TGGCAGCCAAAGGTCGGCGCATCCCTTCAGTGAGATGATGTGCGGAT 1753
835 LeuProPheSerSerIleArgProCysMetLeuAlaThrCysAlaAr 851
1754 CTACAGGGTTCCTCTCTGPAAGATCTTGCAGATGCCGTGAGTCAGTAA 1803
851 g...ProGlyArgProSerThrLys.....HisSerProHisIleA 864
1804 AATCAATCAGAGATGAAGACAAAACCTTGGTGAGCAGGTCGCCAGATCC 1853
864 laAlaAlaArgLysValTyrIleGlnThrArgGlnArgLysLeuHis 880
1854 TCAGTGTCCAGAGAGTACATTACAGCAAGGAAGAGAGCGTATTATAC 1903
881 PheValValGlyGlyPheAlaTyrLeuProLysThrAlaValValLe 897
1904 CTGACCATTTGTAGCAGAGCTATTGTGTCGCCAACACATCCCGTATTAT 1953
897 uArgCysProAlaArgArgValArgLysProLeuIleThrTrpGluLysA 914
1954 TAAGTGCAGTGCAGCATTCAGAAATCTCTGATCCAGTGGGAGAGG 2003
914 spGlyGlnHisLeuLeuSerSerThrHisValThrValAlaProPheGly 930
2004 ATGCCCTTGTCTGCAGAACTCCAAACGGCTTGGCATCCCAAGTCAGGC 2053
931 TyrLeuLysIleHisArgLeuLysProSerAspAlaGlyValTyrThrCy 947
2054 TCATAAAATCCACGGTCTTGTGCTGCCCCACATCGCGTGACCGGTG 2103
947 sSerAlaGlyProAlaArgGluHisPheValIleLysLeuIleGlyA 964
2104 CATTCAGGCTCTGCACAGGAACAGTTGTGCTCAAGCTCATTTGTACTG 2153
964 snArgLysLeuValAlaArgProLeuSerProArgSerGluGluGluVal 980
2154 ACAACGGCTCATCGCAGCCCA..... 2176
981 LeuAlaGlyArgLysGlyGlyProLysGluAlaLeuGlnThrHisLysHi 997
2177 ...GCCCTCAGGAGCCTTATCAGGGAATATCTTGGATGGACCAGCGCA 2223
997 sGlnAsnGlyIlePheSerAsnGlySerLysAlaGluLysArgGlyLeuA 1014
2224 AGCCAATAGTTTGGGAGTCACATGSCACAA.....ATGAGGCAATGT 2267
1014 laAlaAsnProGlySerArgTyr.....AspAspLeuValSerArg... 1027
2268 GGAATACAAAATGACCTTTATCTGGATGATGACCATTTAGTAACAG 2317
1028 .....LeuLeuGluGlnGlyTrpProGlyGluLeuLe 1039
2318 OCTTTCCTGAGAGCTGTGTAGGCCACTGCAGCAATCTGCAGGAAGCAC 2367
1039 uAlaSerTrpGluAlaGlnAspSerAlaGluArgAsnThrThrSerGluG 1056
2368 CAATCTCTGGAGTTGAAGATAAGCAGTTTGAAGCAGCAGTTAAACAA. 2416
1056 LuAspProGlyAlaGluGlnValLeuLeuHisLeuProPheThrMetVal 1072
2417 .....GGAGCA.....TATAGCATGGAT 2434
1073 ThrGluGlnArgArgLeuAspAspIleLeuGlyAsnLeuSerGlnGlnPr 1089
2435 ACAGCCAG.....TTTGTAGCTGATAAGAAACATGATCAGCTCAT 2478
1089 oGlu.....GluLeuArgAspLeuTyrSerLysHisLeuValAlaGlnL 1104
2479 GGAACCGGAGAGGTACCGATGCTTTCGCTCCAGCTGATATATCAGC 2528
1104 euAlaGlnGluIlePheArgSerHisLeuGluHisGlnAspThrLeuLeu 1120

```



```
380 pGluAlaThrProTrpThrAlaCysSerSerCysGlyGlyClyIleG 397
||||| :|||:|||||:|||||:|||||:|||||:|||||:|||||:
1044 GGAACATAATCCCTGGACTGCATGTTCCGTGTCCTGTGGAGGAGGATTC 1093
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
397 InSerArgAlaValSerCysValGluGluAspIleGlnGlyHisValThr 413
|| |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1094 AGACACGAGCTTGTGTGTGTAGAGAAATCCATGTCATGAGAGATATWG 1143
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
414 SerValGluGluThrLysCysMetTyrThrProLysMetProIleAlaG1 430
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1144 CAGTGGGAAGAAATGGAAGTCATGTCATGACCCAAACCCCAAGGTTATGCA 1193
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
430 nProCysAsnIlePheAspCysProLysTyrLysLeuAlaGlnGluTrpSerP 447
| |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1194 AACTGTGTAATCTGTTGATTGCCCAAGTGGATTGCCATGGAGTGGTCTC 1243
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
447 roCysThrValThrCysGlyGlnGlyLeuArgTyrArgValValLeuCys 463
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1244 AGTGACAGTGACTTGTGCGCGAGGGTTACGGTACCGGGTGTCTCTGT 1293
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
464 IleAspHisArgGlyMetHisThrGlyGlyCysSerProLysThrLysPr 480
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1294 ATTAACACCGCGAGAGCATGTGGGGCTGCAATCCAACTGAAGTT 1343
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
480 oHisIleLysGluCysIleValProThrProCysTyrLysProLysG 497
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1344 ACACATCAAGAAGATGTGTCATTCCTCCATCCCGTGTATTAACCAAAAG 1393
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
497 LuLysLeuProValGluAlaLysLeuProThrPheLysGlnAlaGlnGlu 513
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1394 AAAAAAGTCAGTGGAGCAAAATGCGTGGCTGGAACCAAGCAAGAA 1443
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
514 LeuGluGluGlyAlaAlaValSerGluGluProSerPheIleProGluAl 530
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1444 CTAGACAGACCAACAAATAGCAACAGAGAACCAACGTTTCATCCAGAAC 1493
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
530 aTrpSerAlaCysThrValThrCysGlyValGlyThrGlnValArgIleV 547
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1494 CTGTGACGCTGACGTACCAACGCTGTGGCCAGGTGTGCAGGTCCGCGAG 1543
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
547 alArgCysGlnValLeuLeuSerPheSerGlnSerValAlaAspLeuPro 563
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1544 TGAAGTCGGTGTCTCTCTCACATTCACGACGACGTAGACTGAGCTGCC 1593
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
564 IleAspGluCysGluGlyProLysProAlaSerGlnArg..... 576
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1594 GAGGAAGAGTGTGAGGCCCCAGCTGCCACCGACGCGCCTGCTCCT 1643
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
577 ....AlaCysTyrAlaGlyProCysSerGlyGluIleProGluPheAsnP 592
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1644 GGAAGCATGTGATGAGAGCGCGCCTCCGAGAGAGCTAGACATCCCTCTCC 1693
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
592 roAspGluThrAspGlyLeuPheGlyLeuGlnAspPheAspGluLeu 608
|| |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1694 CT.....GAGGACAGTGGAGCAGCT 1713
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
609 TyrAspTrpGluTyrGluGlyPheThrLysCysSerGluSerCysGlyG1 625
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1714 TAGCACTGGGAGTACCGTGGTTCACCCCTGCACACCAACATGCTTGG 1763
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
625 yGlyValGlnGluAlaValValSerCysLeuAsnLysGlnThrArgGluP 642
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1764 AGGCCATCAAGAGCCATACGATGCTGTACATATCCAGACCCAGCAGA 1813
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
642 roAlaGluGluAsnLeuCysValThrSerArgArgProGlnLeuLeu 658
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1814 CAGTCATGACAGCTTGTGTATATGTCACCGCTCTCCAGCCATGAGC 1863
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
659 LysSerCysAsnLeuAspProCysProAlaIleArgTrpGlu 671
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1864 CAGGCTGTAAACAGACAGCGCTGTCCCGCCAGGAGAG 1902
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
```

seq_name: gb_pr:AL591423

seq_documentation_block:

LOCUS AL591423 54193 bp DNA linear PRI 16-NOV-2001

DEFINITION Human DNA sequence from clone RP11-134P18 on chromosome 9, complete sequence.

ACCESSION AL591423

VERSION AL591423.6 GI:16973934

KEYWORDS HTG.

SOURCE human.

ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

AUTHORS Almeida, J.

TITLE Direct Submission

JOURNAL Submitted (16-NOV-2001) Wellcome Trust Sanger Institute, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk

COMMENT On Nov 17, 2001 this sequence version replaced gi:16214807. During sequence assembly data is compared from overlapping clones. Where differences are found these are annotated as variations together with a note of the overlapping clone name. Note that the variation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above. This sequence was finished as follows unless otherwise noted: all regions were either double-stranded or sequenced with an alternate chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one plasmid subclone or more than one M13 subclone; and the assembly was confirmed by restriction digest. The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Em: EMBL; Sw: SWISSPROT; Tr: TREMBL; Wp: WORMPEP; Information on the WORMPEP database can be found at http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence was generated from part of bacterial clone contigs of human chromosome 9, constructed by the Sanger Centre Chromosome 9 Mapping Group. Further information can be found at <http://www.sanger.ac.uk/HGP/Chr9> RP11-134P18 is from the library RPCI-11.1 constructed by the group of Pieter de Jong. For further details see <http://www.chori.org/bacpac/home.htm> VECTOR: pBACE3.6

IMPORTANT: This sequence is not the entire insert of clone RP11-134P18 It may be shorter because we sequenced overlapping sections only once, except for a short overlap. The true left end of clone RP11-220B22 is at 52194 in this sequence. The true right end of clone RP11-503K16 is at 2000 in this sequence.

FEATURES

Location/Qualifiers

source 1..54193

/organism="Homo sapiens"

/db_xref="taxon:9606"

/chromosome="9"

/clone="RP11-134P18"

/clone_lib="RPCI-11.1"

BASE COUNT 15776 a 11257 c 10727 g 16433 t

ORIGIN

alignment_scores:

Quality: 2041.00 Length: 723

Ratio: 4.769 Gaps: 2

Percent Similarity: 59.198 Percent Identity: 58.921

alignment_block:

US-10-044-807-2 x AL591423 ..

Align seg 1/1 to: AL591423 from: 1 to: 54193

799 GluCysSerThrSerCysGlyGluGlyThrGlnThrArgSerAlaIleCy 815
:::|||||
24616 CAGGTTCACCAAGCTCGGGGAAGCCACCAAGCTCGAAGCGCATTTG 24665
815 sArgLysMetLeuLysThrGlyLeuSerThrValValAsnSerThrLeuC 832
|||||
24666 CCGAAGATGCTGAACACCGGCTCTCAACGGTGTCAATTCACCCCTGT 24715
832 ysProProLeuProPheSerSerSerIleArgProCysMetLeuAlaThr 848
|||||
24716 GCGCGCCCTTCCTTCTCTCCATCAGGCCCTGTATGCTGCAACC 24765
849 CysAla..... 850
|||||
24766 TGTGCAAGTAAGTATGTCAGGGCTCTGGGAATGGGGAGATGAACCCACA 24815
850 850
24816 CAGCAGCAGCTAATAGCCACCACGCCGTGGCCTTCCTTAACTCAAGACCT 24865
850 850
24866 GTGGGAATAGTGGACAGTAGAAGCCCATGGAGGCTGCAGGAGGAGA 24915
850 850
24916 GCTCAGCTGGAGACGGAAGGAATCTTAATACACGTGGCCAGGCTAAA 24965
850 850
24966 TTCCTCTCCACTCGGAACACATTTCAGGGGGGTAGACCAAGTTCA 25015
850 850
25016 CTTGCTGGCATACAGAGATCTGGAAGCTGGAGGAGTGTATTACTAATGA 25065
850 850
25066 TGCTAATGTATCCACTCTGGGTGTTTAAACGACTCTAAACAGATTGT 25115
850 850
25116 TCCTGTTAATTAATCCCTAGAGCCGCAACACTTCTAGGCTGTCTTAC 25165
850 850
25166 CATGCTCTTTAGGCTCTGTGCTTGCAGGACCCAGAACCAACATTT 25215
850 850
25216 AGCGGAAGACATCTCTCTCTGTCCTCTGCTTTTATGAGGTTGAT 25265
850 850
25266 TTCTTGTTTAAAAATTATTTAATGAGAGGAATGCTTCCCGGGATC 25315
850 850
25316 CCCTTCCATCTAGAACHCTGCATCTCAGCTTCTAGGACCAACAGGCCCA 25365
850 850
25366 GCAGGGAGACATTGGGTGGCCACTGCATCTCATTTAGTCTGATTGGT 25415
850 850
25416 GGCAGCTTGATGGTCCACACCGGGGAATGAGTCTCTCCCGAACAACAG 25465
850 850
25466 GCAGTAAAGATGCCCCGCCAAAATAAACCCGGAAGATACTTTCTCT 25515
850 850

25516 CCGGTCTCCGGCACCTTGTCTCTCCCTTCTCTTCCACTCTGGCTCTT 25565
850 850
25566 TCCTTTGCCCTCTCTCTGGCTGCATCTCACTGGGTTTTCTCTCTCC 25615
851 Arg...ProGlyA 854
25616 CCACCTCTTTCTCTGCTTCGGTTGCTCTCTCTCCAGGGCCCCGGC 25665
854 rgProSerThrLysHisSerProHisIleAlaAlaAalaArgLysValTyr 870
|||||
25666 GGCCATCCACGAAGCACAGCCGCACATCGCGGCCGCCAGGAAGGTCTAC 25715
871 lleGlnThrArgArgGlnArgLysLeuHisPheValValGlyGlyPheAl 887
|||||
25716 ATACAGACTCGCAGCAGAGAAAGCTGCATCTCGTGGTGGGGGCTTGGC 25765
887 aTyrLeuLeuProLysThrAlaValValLeuArgCysProAlaArgArgV 904
|||||
25766 CTACCTGTCTCCCAAGACGGCGGTGCTGCGCTGCCCGCGCGCAGGG 25815
904 alArgLysProLeuIleThrTrpGluLysAspGlyGlnHisLeuIleSer 920
|||||
25816 TCCGCAAGCCCTCATCACCCTGGGAGAAAGGAGCGCCAGCACCTCATCAGC 25865
921 SerThrHisValThrValAlaProPheGlyTyrLeuLysIleHisArgLe 937
|||||
25866 TCGAGCAGCTCACGGTGGCCCCCTTCGGCTATCTCAAGATCCACCGCT 25915
937 uLysProSerAspAlaGlyValTyrThrCysSerAlaGlyProAlaArgG 954
|||||
25916 CAAGCCCTCGGATCGAGCGCTACACCTGCTCAGCGGGCCCGCCGGG 25965
954 luHisPheValIleLysLeuIleGlyGlyAsnArgLysLeuValAlaArg 970
|||||
25966 AGCACTTTGTATTAAAGCTATCGGAGGCAACCGCAAGCTGCTGGCCCGG 26015
971 ProLeuSerProArgSerGluGluValLeuAlaGlyArgLysGlyG 987
|||||
26016 CCCGTGAGCCCGAGAAGTGAGAGAGAGTGTTCGCGGGGAGGAAGGCGG 26065
987 yProLysGluAlaLeuGlnThrHisLysHisGlnAsnGlyIlePheSerA 1004
|||||
26066 CCCGAGGAGGCCCTGCAGACCCACAAACACAGACGGGATCTTCTCCA 26115
1004 snGlySerLysAlaGluLysArgGlyLeuAlaAlaAsnProGlySerArg 1020
|||||
26116 ACGGCAAGCAGCGGAGAAAGCGGCTTGGCGCCACCCCGGGAGCGGC 26165
1021 TyrAspAspLeuValSerArgLeuLeuGluGlnGlyTyrProGlyG 1037
|||||
26166 TACGACGACCTCTGCTCCCGCTCTGGAGCAGGCGGCTGCGCGGAGA 26215
1037 uLeuLeuAlaSerTrpGluAlaGlnAspSerAlaGluArgAsnThrThrS 1054
|||||
26216 GCTCTGGCTCTGTGGAGGGCGCAGGACTCTGCGGAAGGAAGAACACGACCT 26265
1054 erGluGluAspProGlyAlaGluGlnValLeuLeuHisLeuProPheThr 1070
|||||
26266 CGGAGGAGGACCCGGTGCAGACCAAGTGTCTCTGCACCTGCCCTTACC 26315
1071 MetValThrGluGlnArgArgLeuAspAspIleLeuGlyAsnLeuSerG 1087
|||||
26316 ATGTGTACCGAGCAGCGCGCTGGAGACATCTCTGGGAACCTCTCCCA 26365
1087 nGlnProGluGluLeuArgAspLeuTyrSerLysHisLeuValAlaGlnL 1104
|||||
26366 GCACCCGAGAGCTGCGGACCTCTACAGAACACCTGCTGGTGGCCAGC 26415
1104 euAlaGlnGluIlePheArgSerHisLeuGluHisGlnAspThrLeuLeu 1120
|||||

26416 TGCCGAGGAGATCTTCGAGCCACTGGAGACCAGGACACGCTCCTG 26465
1121 LysProSerGluArgThrSerProValThrLeuSerProHisLysHi 1137
|||||
26466 AAGCCCTGGAGCGCAGGACTTCCCCAGTGACTCTCTCGCCTCATAAACA 26515
1137 sValSerGlyPheSerSerSerLeuArgThrSerSerThrGlyAspAlaG 1154
|||||
26516 CGTGTCTGGCTTCAGCAGCTCCCTGGGACCTCTCCACCGGGGACGCCG 26565
1154 lYgLYGlySerArgArgProHisArgLysProThrIleLeuArgLysIle 1170
|||||
26566 GGGGAGGCTCTCGAAGGCCACACCGCAAGCCACCATCTCTGCCGCAAGATC 26615
1171 SerAlaAlaGlnGlnLeuSerAlaSerGluValValThrHisLeuGlyG1 1187
|||||
26616 TCAGCGGCCCGCAGAGCTCTCAGCCTCGAGGTGTCACCCACCTGGGGCA 26665
1187 nThrValAlaLeuAlaSerGlyThrLeuSerValLeuLeuHisCysGluA 1204
|||||
26666 GACGGTGGCCCTGGCCAGCGGACACTGAGTGTCTTCTGCACTGTGAGG 26715
1204 lalLeGlyHisProArgProThrIleSerTrpAlaArgAsnGlyGluGlu 1220
|||||
26716 CCATCGGCCACCCCAAGGCCCTACCATCAGCTGGGCCAGGAATGGAGAAGAA 26765
1221 ValGlnPheSerAspArg 1226
|||||
26766 GTTCAGTTCAGTGACAGG 26783

OM of: US-10-044-807-2 to: N_Geneseq_032802.* out_format : pfs

Date: Jul 24, 2002 6:44 AM

About: Results were produced by the GenCore software, version 4.5,
Copyright (c) 1993-2000 CompuGen Ltd.

Command line parameters:

-MODEL=frame+p2n_pool -DEV=xlh
-O=/cgn2_1/USPTO_spool/US10044807/runat_22072002_153742_13701/app_query.fasta_1.1847
-DB=N_Geneseq_032802 -QFMT=fastap -SUFFIX=ring -GAPOP=12.000
-GAPEXT=4.000 -MINMATCH=0.100 -LOOPL=0.000 -LOOPEXT=0.000
-GAPOP=4.500 -GAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500
-FEAPOP=6.000 -DELEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500
-DELOP=6.000 -DELEXT=7.000 -START=1 -MATRIX=blossum62
-TRANS=human40.cdi -LIST=45 -DOCLALIGN=200 -THR_SCORE=pct
-THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=pfs
-NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US10044807 -CGN1_1.264 -NCPU=6 -ICPU=3 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -NO_XLPAY -WAIT -THREADS=1

Search information block:

Query: US-10-044-807-2
Query length: 1762
Database: N_Geneseq_032802.*
Database sequences: 1736436
Database length: 856457221
Search time (sec): 392.130000

score_list:

Sequence	Strd Orig	Zscore	EScore	Len	Documentation
/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA2001A.DAT:AAF97891			9466.50	8712.10	0.0
/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA2001A.DAT:AAH9417			9464.50	8710.18	0.0
/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA2001A.DAT:AAF97892			6176.50	5682.89	3.4e-3
/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA2000.DAT:AAH47482			4797.50	4413.94	1.6e-23
/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA2000.DAT:AAH47458			4775.00	4393.20	2.3e-23
/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA2001A.DAT:AAH47458			3956.50	3632.64	5.4e-11
/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA2001A.DAT:AAH47757			3953.50	3626.58	1.2e-11
/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA2001A.DAT:AAH47371			3762.50	3454.17	4.7e-11
/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA2001A.DAT:AAH47372			3673.00	3366.41	3.7e-11
/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA2000.DAT:AAH98232			3612.00	3319.69	1.5e-17
/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA2001A.DAT:AAH98232			3423.00	3149.15	4.6e-11
/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1999.DAT:AAH34146			2919.00	2685.19	3.2e-14
/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA2000.DAT:AAH78532			2919.00	2685.19	3.2e-14
/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA2001A.DAT:AAH46142			2919.00	2685.19	3.2e-14
/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA2001A.DAT:AAH63447			2910.00	2677.23	8.9e-11
/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA2001A.DAT:AAH4369			2809.50	2581.41	1.9e-11
/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA2001A.DAT:AAH4370			2615.50	2403.38	1.6e-11
/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA2001A.DAT:AAH7756			2597.50	2377.57	4.4e-11
/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA2001A.DAT:AAH7758			2558.00	2353.67	9.4e-11
/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA2001A.DAT:AAH66429			2425.50	2228.76	8.5e-11
/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA2000.DAT:AAH75770			2351.00	2164.81	3.1e-11
/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA2001B.DAT:AAH81125			2346.50	2153.23	1.4e-11
/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA2001A.DAT:AAH4368			2231.50	2050.87	6.9e-11
/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA2001A.DAT:AAH08462			1930.00	1778.09	1.1e-9
/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA2000.DAT:AAH74709			1796.50	1640.51	5.0e-83
/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA2001A.DAT:AAH4365			1525.00	1402.90	8.5e-7
/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1999.DAT:AAH97684			1476.00	1360.45	2.0e-67
/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA2001A.DAT:AAH97927			1410.00	1301.03	4.0e-6
/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1999.DAT:AAH97583			1341.00	1238.62	1.2e-60
/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA2001A.DAT:AAH4366			1331.00	1225.68	6.3e-6
/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA2000.DAT:AAH36862			1151.50	1043.76	8.6e-50
/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA2001A.DAT:AAH08559			1111.00	1025.40	9.1e-4
/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA2002.DAT:AAH97179			1004.00	909.01	2.8e-42
/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA2002.DAT:AAH97180			999.00	906.71	3.7e-42
/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA2001A.DAT:AAH63449			991.50	887.20	1.3e-4
/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA2001A.DAT:AAH63443			975.00	882.28	8.5e-4
/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA2001B.DAT:AAH92052			947.00	872.52	3.0e-4
/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA2000.DAT:AAH75588			947.00	855.41	2.7e-39
/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA2001B.DAT:ABL02495			941.00	846.60	8.3e-3

/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA2001A.DAT:AAF82149 + 887.00 802.40 2.4
/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA2001A.DAT:AAH9417 + 886.00 817.77 3.3
/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA2001A.DAT:AAH04475 + 886.00 800.96 2.9
/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA2001A.DAT:AAH49372 + 885.00 800.06 3.2
/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA2001A.DAT:AAH04494 + 885.00 800.04 3.2

seq_name: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA2001A.DAT:AAF97891

seq_documentation_block:

ID AAF97891 standard; cDNA; 5720 BP.

XX AAF97891;

XX 01-JUN-2001 (first entry)

XX Human secreted protein cDNA, SEQ ID NO: 18.

XX Human; secreted protein; immunomodulatory; antisclerotic;

KW dermatological; antiinflammatory; anti-HIV; cytostatic; cardiant;

KW vascular; anti-angiogenic; ophthalmological; neuroprotectant;

KW neotropic; anticonvulsant; antialzheimers; antiparkinsonian;

KW antimicrobial; vulnary; vaccine; gene therapy; cancer;

KW protein coordinate data; infection; ss.

XX Homo sapiens.

XX WO200121658-A1. *Pindry*

XX 29-MAR-2001.

XX 22-SEP-2000; 2000WO-US26013.

XX 24-SEP-1999; 99US-0155709.

XX (HUMA-) HUMAN GENOME SCI INC.

PI Ni J, Baker KP, Birse CE, Ebner R, Fiscella M, Komatsoulis GA;

PI Laflaur DW, Moore PA, Olsen HS, Rosen CA, Ruben SA, Soppet DR;

PI Young PE, Wei P, Florence KA;

DR WPI; 2001-235311/24.

XX Nucleic acids encoding 32 human secreted polypeptides, useful for

PT preventing, diagnosing and/or treating e.g. cancers, Parkinson's

PT disease and diabetic retinopathy -

PS Claim 1; Fig 4; 890pp; English.

CC The present sequence encodes one of 32 novel human secreted polypeptides.

CC The nucleic acid molecules and polypeptides they encode may be used in

CC the prevention, diagnosis and treatment of diseases such as

CC immune disorders (e.g. multiple sclerosis, systemic lupus erythematosus

CC and human immuno-deficiency virus (HIV) infections), hyperproliferative

CC disorders (e.g. cancers and Gaucher's disease), cardiovascular diseases

CC (e.g. Scimitar syndrome, Chaga's cardiomyopathy and coronary

CC arteriosclerosis), angiogenic disorders (e.g. corneal graft

CC neovascularisation and diabetic retinopathy), neurological disorders

CC (e.g. Huntington's chorea, Alzheimer's disease and Parkinson's disease),

CC infectious diseases and/or for promoting wound healing, regeneration

CC and/or chemotaxis. The nucleic acid molecules may be used to produce the

CC secreted polypeptides. They may also be used as DNA probes in diagnostic

CC assays to detect and quantify the presence of similar nucleic acid

CC sequences in samples. The polypeptides may be used as antigens in the

CC production of antibodies and in assays to identify modulators of

CC their expression and activity.

XX Sequence 5720 BP; 1293 A; 1677 C; 1611 G; 1139 T; 0 other;

alignment_scores:

Quality: 9466.50 Length: 1762

Ratio: 5.425 Gaps: 1

Percent Similarity: 99.035 Percent Identity: 98.978

alignment_block:

US-10-044-807-2 x AAF97891 ..

Align seg 1/1 to: AAF97891 from: 1 to: 5720

1 MetGluCysCysArgAlaThrProGlyThrLeuLeuLeuPheLeuAl 17
|||||
67 ATGGAATGCTGCCGTCGGCAACTCCCTGGCACACTGCTCTCTTTCTGGC 116
17 aPheLeuLeuLeuSerSerArgThrAlaAaSerGluCluAspArgAspG 34
|||||
117 TTCTCCTGCTCCTGAGTTCCAGAGCCGACGCTCCGAGGAGAGCCGGGACG 166
34 lyLeuTrpAspAlaTrpGlyProThrSerGluCysSerArgThrCysGly 50
|||||
167 GCCTATGGAGTCCCTGGGGCCCATGGAGTGAATGCTCACGCACTCGCGG 216
51 GlyGlyAlaSerTyrSerLeuArgArgCysLeuSerSerLysSerCysGl 67
|||||
217 GGTGGGGCCTCCTACTCTGTAGGCGCTGCTGAGCAGCAAGAGCTGTGA 266
67 uGlyArgAsnIleArgTyrArgThrCysSerAsnValAspCysProProG 84
|||||
267 AGGAAGAAATATCCGATACAGACATCGATTAATGTGGACTGCCACAG 316
84 luAlaGlyAspPheArgAlaGlnGlnCysSerAlaHisAsnAspValLys 100
|||||
317 AAGCAGGTGATTTCCGAGCTCAGCAATGCTCAGCTCATATGATGTCAG 366
101 HisHisGlyGlnPheTyrGluTrpLeuProValSerAsnAspProAspAs 117
|||||
367 CACCATGGCCAGTTTATGAATGGCTTCCTGTGTCTAATGACCCCTGACAA 416
117 nProCysSerLeuLysCysGlnAlaLysGlyThrThrLeuValValGluL 134
|||||
417 CCCATGTTCACTCAAGTCCCAAGCCCAAGCAACACCTGGTGTGTGAC 466
134 euAlaProLysValLeuAspGlyThrArgCysTyrThrGluSerLeuAsp 150
|||||
467 TAGCACCTAAGTCTTAGATGGTACGGCTTGTCTATACAGATCTTTGGAT 516
151 MetCysIleSerGlyLeuCysGlnIleValGlyCysAspHisGlnLeuGl 167
|||||
517 ATGTGCATCAGTGTGTATGCCAAATTTGTTGGCTCGCATCCAGCTGGG 566
167 ySerThrValLysGluAspAsnCysGlyValCysAsnGlyAspGlySerT 184
|||||
567 AAGCACCGCTCAAGAGAGTAACTGTGGGGCTCGCAACGGAGATGGGTCCA 616
184 hrCysArgLeuValArgGlyGlnTyrLysSerGlnLeuSerAlaThrLys 200
|||||
617 CCTGCCGCTGGTCCGAGGGCAGTATAAATCCCAAGCTCTCGGCAACCAAA 666
201 SerAspAspThrValValAlaIleProTyrGlySerArgHisIleArgLe 217
|||||
667 TCGGATGATACGTGGTGTGCAATTCCTTATGGAAGTAGACATATTCGGCT 716
217 uValLeuLysGlyProAspHisLeuTyrLeuGluThrLysThrLeuGlnG 234
|||||
717 TGTCTTAAAGGTCTCTGATCATCTTATATCTGGAACCAACCAACCTCCAGG 766
234 lyThrLysGlyCluAsnSerLeuSerSerThrGlyThrPheLeuValAsp 250
|||||
767 GGACTAAAGGTGAACAACAGTCTCAGCTCCACAGGAACCTTCCTTGTGGAC 816
251 AsnSerSerValAspPheGlnLysPheProAspLysGluIleLeuArgMe 267
|||||
817 AATTCTAGTGGGACTTCCAGAAATTTCCACACAAAGAGATACAGAGAT 866
267 tAlaGlyProLeuThrAlaAspPheIleValLysIleArgAsnSerGlyS 284
|||||
867 GGCTGGACCACTCACAGCAGATTTTCATTTGTCAGATTCGTAACCTCGGCT 916

284 erAlaAspSerThrValGlnPheIlePheTyrGlnProIleIleHisArg 300
|||||
917 CCGCTGCAGTACAGTCCAGTTCTTCTATCAACCCATCATCCACCGA 966
301 TrpArgGluThrAspPhePheProCysSerAlaThrCysGlyGlyTy 317
|||||
967 TGGAGGGAGACGGATTTCTTCTTGTCTCAGCACTGTGGAGAGGTTA 1016
317 rGlnLeuThrSerAlaGluCysTyrAspLeuArgSerAsnArgValVala 334
|||||
1017 TCAGCTGACATCGGCTGAGTGTCTGAGTCTGAGGAGCAACCGTGGTGG 1066
334 laAspGlnTyrCysHisTyrTyrProGluAsnIleLysProLysProLys 350
|||||
1067 CTGACCAATACTGTCTCATTATACCCAGAGAACATCAAAACCCAAACCAAG 1116
351 LeuGlnGluCysAsnLeuAspProCysProAlaSerAspGlyTyrLysGl 367
|||||
1117 CTTCAGGAGTGCACCTGGATCCTTGTCCAGCC..... 1149
367 nIleMetProTyrAspLeuTyrHisProLeuProArgTyrGluAlaThrP 384
|||||
1150AGGTGGAGGCCACCC 1165
384 roTrpThrAlaCysSerSerCysGlyGlyIleGlnSerArgAla 400
|||||
1166 CATGGACCGCTGCTCCTCCTGTGTGGGGGGCATCCAGAGCCGGCA 1215
401 ValSerCysValGluAspIleGlnGlyHisValThrSerValGluGl 417
|||||
1216 GTTTCCTGTGGAGGAGACATCCAGGGGCATGTCATTCTCAGTGAAGA 1265
417 uTrpLysCysMetTyrThrProLysMetProIleAlaGlnProCysAsnI 434
|||||
1266 GTGGAAATGCATGTACACCCCTAAGATGCCATCGCGCAGCCCTGCAACA 1315
434 lePheAspCysProLysTrpLeuAlaGlnGluTrpSerProCysThrVal 450
|||||
1316 TTTTGTGACTGCCCTAAATGGCTGGCACAGGAGTGGTCTCCGTGCACAGTG 1365
451 ThrCysGlyGlnGlyLeuArgTyrArgValValLeuCysIleAspHisAr 467
|||||
1366 ACATGTGGCAGGGCCTCAGATACCGTGTGGTCTCTGATCGACCATCG 1415
467 gGlyMetHisThrGlyGlyCysSerProLysThrLysProHisIleLysG 484
|||||
1416 AGGAATGCACACAGGAGGCTGTAGCCCAAAACAAAGCCCAACATAAAG 1465
484 luGluCysIleValProThrProCysTyrLysProLysGluLysLeuPro 500
|||||
1466 AGGAATGCATCGTACCCACTCCTCTCTATAAACCCCAAGAGAACTTCCA 1515
501 ValGluAlaLysLeuProTrpPheLysGlnAlaGlnGluLeuGluGluGl 517
|||||
1516 GTGAGGCCAAGTTCCTCATGGTTCAAACAAGCTCAAGAGCTAGAAGAAG 1565
517 yAlaAlaValSerGluProSerPheIleProGluAlaTrpSerAlaC 534
|||||
1566 AGCTGCTGTGCAGAGGAGCCCTGTTTCATCCAAAGGCTGTGGCT 1615
534 ysThrValThrCysGlyValGlyThrGlnValArgIleValArgCysGln 550
|||||
1616 GCACAGTCACTGTGGTGTGGGGACCCAGGTGCGAATAGTCAGGTGCCAG 1665
551 ValLeuLeuSerPheSerGlnSerValAlaAspLeuProIleAspGluCy 567
|||||
1666 GTGCTCCTGTCTTCTCTCAGTCCGCTGGCTGACCTGCCTATTGACGAGTG 1715
567 sGluGlyProLysProAlaSerGlnArgAlaCysTyrAlaGlyProCysS 584
|||||
1716 TGAAGGGCCCAAGCCAGCATCCAGCGTGCCTGCTTATGTCAGGCCCATGCA 1765

584 erGlyGluIleProGluPheAsnProAspGluThrAspGlyLeuPheGly 600
|||||
1766 GGGGGGAATTCTGAGTTCAACCCAGACAGACAGATGGGCTCTTTGGT 1815
601 GlyLeuGlnAspPheAspGluLeuTyrAspTrpGluTyrGluGlyPheTh 617
|||||
1816 GCCTTCACAGATTTCACACAGCTGATGACTGGGAGTATGAGGGGTTCAC 1865
617 rLysCysSerGluSerCysGlyGlyValGlnGluAlaValValSerC 634
|||||
1866 CAAGTGCTCCGAGTCTGTGGAGGAGGTCTCCAGGAGCTGTGGTGAAGCT 1915
634 yLsLeuAsnLysGlnThrArgGluProAlaGluAlaAsnLysValThr 650
|||||
1916 GCTTGAACAACACAGACTCGGAGCCTGCTGAGAGAACCTGTGCGTGACC 1965
651 SerArgArgProGlnLeuLeuLysSerCysAsnLeuAspProCysPr 667
|||||
1966 AGCCGCGCGCCGCCACAGCTCCTGAAAGTCTGCAATTGGATCCCTGSCC 2015
667 oAlaArgTTPGluIleGlyLysTrpSerProCysSerLeuThrCysGlyV 684
|||||
2016 AGCAAGGTGGGAATTGGCAAGTGGAGTCCATGTAGTCTCACATGTGGG 2065
684 aLcLysLeuGlnThrArgAspValPheCysSerHisLeuLeuSerArgGlu 700
|||||
2066 TCGGCTACAGACCAGAGACGTCTTCTGCAGCCACCTGCTTTCCAGAGAG 2115
701 MetAsnGluThrValIleLeuAlaAspGluLeuCysArgGlnProLysPr 717
|||||
2116 ATGAATGAACAACAGTATCTCGGTGATGAGTGTGTGCGCAGCCCAAGCC 2165
717 oSerThrValGlnAlaCysAsnArgPheAsnCysProProAlaTrpTyrP 734
|||||
2166 CAGCAGCGTGCAAGCTGTGAACCGCTTTAATGCCCCCCAGCCGTGTACC 2215
734 roAlaGlnTrpGlnProCysSerArgThrCysGlyGlyValGlnLys 750
|||||
2216 CTGCACAGTGGCAGCGCTGTTCAGAACCTGTGGCGGGGTGTTCAGAA 2265
751 ArgGluValLeuCysLysGlnArgMetAlaAspGlySerPheLeuGluLe 767
|||||
2266 CGTGAAGTCTTTGCAAGCAGCGCATGGGTGATGGCAGCTTCTCGAGCT 2315
767 uProGluThrPheCysSerAlaSerLysProAlaCysGlnGlnAlaCysL 784
|||||
2316 TCTTGAGACTTCTGTTCAGCTTCANAACTGCTGCCAGCAAGCATGCA 2365
784 yLysAspCysProSerGluTrpLeuLeuSerAspTrpThrGluCys 800
|||||
2366 AGAAAGATGACTTCCACAGCGAGTGGCTTCTCTCAGACTGGACAGATGT 2415
801 SerThrSerCysGlyGluGlyThrGlnThrArgSerAlaIleCysArgLy 817
|||||
2416 TCACAAGCTGCGGGGAAGCACCACAGACTCGAAGCGCCCATTTGCCGAAA 2465
817 sMetLeuLysThrGlyLeuSerThrValValAsnSerThrLeuCysProp 834
|||||
2466 GATGCTGAAACCGGCTCTCAACGGTTGTCAATTCACCCCTGTGCCCGC 2515
834 roLeuProPheSerSerIleArgProCysMetLeuAlaThrCysAla 850
|||||
2516 CCCTGCTTCTCTCTCCATCAGGCCCTGTATGCTGGCAAGCTGTGCA 2565
851 ArgProGlyArgProSerThrLysHisSerProHisIleAlaAlaAr 867
|||||
2566 AGGCCGCGCGGCCATCCAGAAAGCAGCGCCGCACATCCGCGCGCCAG 2615
867 qLysValTyrIleGlnThrArgArgGlnArgLysLeuHisPheValValG 884
|||||
2616 GAAGGTCTACATACAGACTCGCAGGACAGAGAGAGTGCACCTTCGTGGTG 2665
884 lyGlyPheAlaTyrLeuLeuProLysThrAlaValValLeuArgCysPro 900
|||||

2666 GGGGCTTCGCTACCTGCTCCCCAAGACGGCGTGGTGTGCTGCCCTGCCCG 2715
901 AlaArgValArgLysProLeuIleThrTrpGluLysAspGlyGlnHi 917
|||||
2716 GCGCGAGGGTCCGCAAGCCCTCATCAGCTGGGAAAGACGCCAGCA 2765
917 steuIleSerSerThrHisValThrValAlaProPheGlyTyrLeuLysI 934
|||||
2766 CCTCATCAGCTCGACGACGTCACGTTGGGCCCTTCGGCTATCTCAGA 2815
934 leHisArgLeuLysProSerAspAlaGlyValTyrThrCysSerAlaGly 950
|||||
2816 TCCACCGCTCAAGCCCTCGGATCGAGCGCTTACACCTGCTCAGCGGC 2865
951 ProAlaArgGluHisPheValIleLysLeuIleGlyGlyAsnArgLysLe 967
|||||
2866 CCGGCCCGGAGACATTTGTGATTAAGTCTATCGAGGCAACGCCAAGCT 2915
967 uValAlaArgProLeuSerProArgSerGluGluValLeuAlaGlyA 984
|||||
2916 CGTGGCCCGCCCTTGAGCCCGAGAAGTGAAGAGGTGCTTTCGGGGA 2965
984 rGlyGlyGlyProLysGluAlaLeuGlnThrHisLysHisGlnAsnGly 1000
|||||
2966 GGAAGGCGCGCCGGAAGAGCGCTCGAGACCCACAAACACAGAACGG 3015
1001 IlePheSerAsnGlySerLysAlaGluLysArgGlyLeuAlaAlaAsnPr 1017
|||||
3016 ATCTTCTCCAAACGGCAGCAAGCGGAGAAAGCGGGGCTTGGCCGCCAACCC 3065
1017 oGlySerArgTyrAspLeuValSerArgLeuLeuGluGlnGlyGlyT 1034
|||||
3066 GGGAGCCGCTACGACGACCTCGTCTCCCGGCTGCTGGAGCAGGGCGGCT 3115
1034 rpProGlyGluLeuLeuAlaSerTrpGluAlaGlnAspSerAlaGluArg 1050
|||||
3116 GGCCCGGAGAGCTGCTGGCCCTCGTGGAGGCGCAGGACTCTCGCGAAAGG 3165
1051 AsnThrThrSerGluGluAspProGlyAlaGluGlnValLeuLeuHisLe 1067
|||||
3166 AACAGACCTCGGAGGAGGACCCGGGTGCAGAGCAAGTGTCTCTGCCACT 3215
1067 uProPheThrMetValThrGluGlnArgArgLeuAspAspIleLeuGlyA 1084
|||||
3216 GCCCTTCACCATGCTGACCGCAGCAGCGCGCTGGACGACATCCTGSGGA 3265
1084 snLeuSerGlnGlnProGluGluLeuArgAspLeuTyrSerLysHisLeu 1100
|||||
3266 ACCTCTCCACAGCCCGAGAGCTGCGCACCTCTACAGCAAGCACCTG 3315
1101 ValAlaGlnLeuAlaGlnGluIlePheArgSerHisLeuGluHisGlnAs 1117
|||||
3316 GTGCCAGCTGGCCAGGAGATCTCCGAGCCACCTGGAGCACCAGGA 3365
1117 pThrLeuLeuLysProSerGluArgArgThrSerProValThrLeuSerP 1134
|||||
3366 CACGCTCTGAAGCCCTCGGAGCGCAGGACTTCCCAGTGACTCTCTCGC 3415
1134 roHisLysHisValSerGlyPheSerSerSerLeuArgThrSerSerThr 1150
|||||
3416 CTCAATAACACGTCTCTGGCTTACGACGCTCCCTCGCGAGCTCTCTCCACC 3465
1151 GlyAspAlaGlyGlySerArgArgProHisArgLysProThrIleLe 1167
|||||
3466 GGGGACCCCGGGGAGGCTCTCGAAGGCCACACCGCAAGCCGCCATCCT 3515
1167 uArgLysIleSerAlaAlaGlnGlnLeuSerAlaSerGluValValThrH 1184
|||||
3516 GCACAAGATCTCAGCGGCCAGCAGCTCTCAGCCTCGGAGGTGTCTACCC 3565
1184 isLeuGlyGlnThrValAlaLeuAlaSerGlyThrLeuSerValLeuLeu 1200
|||||

3566 ACCTGGGCAGACGGTGGCCCTGGCCAGCGGGACACTGAGTGTCTTCTCTG 3615
1201 HisCysGluAlaIleGlyHisProArgProThrIleSerTrpAlaArgAs 1217
|||||
3616 CACTGTGAGGCCATCGGCCACCCAAAGCCCTACCATCAGCTGGGCCAGAA 3665
1217 nGlyGluValGlnPheSerAspArgIleLeuLeuGlnProAspAsps 1234
|||||
3666 TGGAGAAAGATTTCAGTTCAGTGCACAGGATTCTTACAGCCAGATGATT 3715
1234 erLeuGlnIleLeuAlaProValGluAlaAspValGlyPheTyrThrCys 1250
|||||
3716 CCTACAGATCTTGGCCACAGTGGAGCAGATGTGGGTTCACACTTGC 3765
1251 AsnAlaThrAsnAlaLeuGlyTyrAspSerValSerIleAlaValThrIe 1267
|||||
3766 AATGCCACCAATGCCITGGCATACGACTCTGTCTCCATTGCCGTCACATT 3815
1267 uAlaGlyLysProLeuValLysThrSerArgMetThrValIleAsnThrG 1284
|||||
3816 AGCAGAAAGCCCACTAGTGAACACGTCACGAATGACAGTGATCAACACGG 3865
1284 luiYProAlaValThrValAspIleGlySerThrIleLysThrValGln 1300
|||||
3866 AGAAGCCTGCAGTCACAGTCGATATAGGAGCACCACATCAAAACAGTGCAG 3915
1301 GlyValAsnValThrIleAsnCysGlnValAlaGlyValProGluAlaG 1317
3916 GGAGTGAATGTGACAATCAACTGCCAGGTTCAGAGTGCCTGAAGCTGA 3965
1317 uValThrTrpPheArgAsnLysSerLysLeuGlySerProHisHisLeuH 1334
|||||
3966 AGTCACCTTGGTTCAGGAATAAAGCAAACTGGGGCTCCCGCACCATCTGC 4015
1334 isGluGlySerLeuLeuLeuThrAsnValSerSerSerAspGlnGlyLeu 1350
|||||
4016 ACGAAGCTCCTTGTCTGCACAAACGTGTCTCTCGGATCAGGSCCTG 4065
1351 TyrSerCysArgAlaAlaAsnLeuHisGlyLeuThrGluSerThrG 1367
4066 TACTCTGGAGGGCGGCCAATCTTCATGGAGAGCTGACTGAGAGCACCCA 4115
1367 nLeuLeuIleLeuAspProGlnValProThrGlnLeuGluAspIleA 1384
|||||
4116 GCTGTGATCTTAGATCCCTCCCAAGTCCCCACACAGTGGAAAGACATCA 4165
1384 rGAlaLeuLeuAlaIaThrGlyProAsnLeuProSerValLeuThrSer 1400
|||||
4166 GGGCCTTGTCTGCTGCCACTGGACCGCAACCTCTCTCAGTGTGAGCTCT 4215
1401 ProLeuGlyThrGlnLeuValLeuAspProGlyAsnSerAlaLeuLeuG 1417
|||||
4216 CCTCTGGGAACACAGCTGTGTCTGGATCTGGGAATTCGTCTCTCTCTGG 4265
1417 yCysProIleLysGlyHisProValProAsnIleThrTrpPheHisGlyG 1434
|||||
4266 CTGCCCATCAAGGTCACCTGTCCCTAATATCACCTGTTTCATGGTG 4315
1434 lYcInProIleValThrAlaThrGlyLeuThrHisHisIleLeuAlaIa 1450
4316 GTCAGCAATTTGCTATGTCACAGGACTGACGCATCACAATCTTGGCAGCT 4365
1451 GlyGlnIleLeuGlnValAlaAsnLeuSerGlySerClnGlyGluPh 1467
4366 GGACATCCTTCAAGTTCGAACCTTACCGTGGGTCTCAAGGGGAATT 4415
1467 eSerCysLeuAlaGlnAsnGluAlaGlyValLeuMetGlnLysAlaSerL 1484
|||||
4416 CAGCTGCTTGTCTCAGAAATGAGCAGGGGTGCTCATGACAGAGGCATCTT 4465
1484 euValIleGlnAspTyrTrpTrpSerValAspArgLeuAlaThrCysSer 1500
|||||
4466 TAGTGATCCAAAGATTACTGGTGGTGTGTGGACAGACTGGCAACCTGCTCA 4515

1501 AlaSerCysGlyAsnArgGlyValGlnGlnProArgLeuArgCysLeuLe 1517
|||||
4516 GCCTTCCTGTGTAAACCGGGGGTTTCAGACGCCCGCTTGAGGTCCCTGCT 4565
1517 uAsnSerThrGluValAsnProAlaHisCysAlaGlyLysValArgProA 1534
|||||
4566 GAACAGCACGAGGTCAACCTGCCACTGCGCAGGGAAGGTTCCGCCCTG 4615
1534 laValGlnProIleAlaCysAsnArgAspCysProSerArgTrpMet 1550
|||||
4616 CGGTGAGCCCCATCGCTGCAACCGAGAGACTGCCCTTCTCGTGGATG 4665
1551 ValThrSerTrpSerAlaCysThrArgSerCysGlyGlyGlyValGlnTh 1567
|||||
4666 GTGACCTCCTGGTCTGCTGTATCCCGGAGCTGTGGGGAGGTGTCCAGAC 4715
1567 rArgArgValThrCysGlnLysLeuLysAlaSerGlyIleSerThrProv 1584
|||||
4716 CCGCAGGCTGACCTGTCAAAAGCTGAAAGGCTCTGGGATCTCCACCGCTG 4765
1584 alSerAsnAspMetCysThrGlnValAlaLysArgProValAspThrGln 1600
|||||
4766 TGTCCAATGACATGTGACCCAGGTGCGCAAGCGGCTGTGGACACCCAG 4815
1601 AlaCysAsnGlnGlnLeuCysValGluTrpAlaPheSerSerTrpGlyG 1617
|||||
4816 GCCTGTAAACCAAGCAGCTGTGTGTGAGCTGGGCTTCTCCAGCTGGGGCCA 4865
1617 nCysAsnGlyProCysIleGlyProHisLeuAlaValGlnHisArgGlnV 1634
4866 GTGCATGGGCTTGCATCGGCTCACCTAGCTGTGCAACACACAGACAAG 4915
1634 alPheCysGlnThrArgAspGlyIleThrLeuProSerGluGlnCysSer 1650
|||||
4916 TCTTCTGCCAGACACGGGATGGCATCACCTTACCATCAGACGAGTGAGT 4965
1651 AlaLeuProArgProValSerThrGlnAsnCysTrpSerGluAlaCysSe 1667
|||||
4966 GCTCTTCCGAGGCTGTGAGCACCAGAACCTGCTGGTCCAGAGCCTCGAG 5015
1667 rValHisTrpArgValSerLeuTrpThrLeuCysThrAlaThrCysGlyA 1684
5016 TGTACACTGGAGAGTCACTGCTGGACCTGTGCACAGCTACCTGTGGCA 5065
1684 snTyrGlyPheGlnSerArgValGluCysValHisAlaArgThrAsn 1700
|||||
5066 ACTACGGCTTCCAGTCCCGCGGTGTGGAGTGTGTGCATGCCGACCAAC 5115
1701 LysAlaValProGluHisLeuCysSerTrpGlyProArgProAlaAsnTr 1717
|||||
5116 AAGGCAGTGCCTGAGCACCTGTCTCTCTGGGGCCCGGCTGCCAACTG 5165
1717 pGlnArgCysAsnIleThrProCysGluAsnMetGluCysArgAspThrT 1734
5166 GCAGCGCTGCACATCACCCCATGTGAAACATGGAGTGGAGTGAGAGACCA 5215
1734 hrArgTyrCysGluLysValLysGlnLeuLysLeuCysGlnLeuSerGln 1750
5216 CCAGGTACTGCGAGAGAGGTGAACAGCTGAACCTCTGCCAACTCAGCCAG 5265
1751 PheLysSerArgCysCysGlyThrCysGlyLysAla 1762
5266 TTTTAAATCTCGTCTGTGGAACTGTGGCAAAAGCG 5301

seq_name: /SIDSL/gcgdata/hold- geneseq/geneseq-emb1/NA2001A.DAT:ABA06468

seq_documentation_block:

ID ABA06468 standard; cDNA; 5769 BP.

XX

ABA06468;

XX

DT 10-JAN-2002 (first entry)

XX Human cDNA SEQ ID NO: 134.
XX Human; gene therapy; neural disorder; immune system disorder;
KW muscular disorder; reproductive disorder; gastrointestinal disorder;
KW pulmonary disorder; cardiovascular disorder; renal disorder;
XX proliferative disorder; inflammation; ss.
XX
OS Homo sapiens.
XX
XX WO200154474-A2. *Prubing*
XX
XX 02-AUG-2001.
XX
XX 17-JAN-2001; 2001WO-US01349.
XX
XX 31-JAN-2000; 2000US-179065P.
XX 04-FEB-2000; 2000US-180828P.
XX 24-FEB-2000; 2000US-184664P.
XX 02-MAR-2000; 2000US-186350P.
XX 16-MAR-2000; 2000US-189874P.
XX 17-MAR-2000; 2000US-190076P.
XX 18-APR-2000; 2000US-198123P.
XX 19-MAY-2000; 2000US-205515P.
XX 07-JUN-2000; 2000US-209467P.
XX 28-JUN-2000; 2000US-214886P.
XX 30-JUN-2000; 2000US-215135P.
XX 07-JUL-2000; 2000US-216647P.
XX 07-JUL-2000; 2000US-216880P.
XX 11-JUL-2000; 2000US-217487P.
XX 11-JUL-2000; 2000US-217496P.
XX 14-JUL-2000; 2000US-218290P.
XX 26-JUL-2000; 2000US-220963P.
XX 26-JUL-2000; 2000US-220964P.
XX 14-AUG-2000; 2000US-224518P.
XX 14-AUG-2000; 2000US-224519P.
XX 14-AUG-2000; 2000US-225113P.
XX 14-AUG-2000; 2000US-225214P.
XX 14-AUG-2000; 2000US-225266P.
XX 14-AUG-2000; 2000US-225267P.
XX 14-AUG-2000; 2000US-225268P.
XX 14-AUG-2000; 2000US-225270P.
XX 14-AUG-2000; 2000US-225447P.
XX 14-AUG-2000; 2000US-225757P.
XX 14-AUG-2000; 2000US-225758P.
XX 14-AUG-2000; 2000US-225759P.
XX 18-AUG-2000; 2000US-226279P.
XX 22-AUG-2000; 2000US-226681P.
XX 22-AUG-2000; 2000US-226868P.
XX 22-AUG-2000; 2000US-227182P.
XX 23-AUG-2000; 2000US-227009P.
XX 30-AUG-2000; 2000US-228924P.
XX 01-SEP-2000; 2000US-229287P.
XX 01-SEP-2000; 2000US-229343P.
XX 01-SEP-2000; 2000US-229344P.
XX 01-SEP-2000; 2000US-229345P.
XX 05-SEP-2000; 2000US-229509P.
XX 05-SEP-2000; 2000US-229513P.
XX 06-SEP-2000; 2000US-230437P.
XX 06-SEP-2000; 2000US-230438P.
XX 08-SEP-2000; 2000US-231242P.
XX 08-SEP-2000; 2000US-231243P.
XX 08-SEP-2000; 2000US-231244P.
XX 08-SEP-2000; 2000US-231413P.
XX 08-SEP-2000; 2000US-231414P.
XX 08-SEP-2000; 2000US-232080P.
XX 08-SEP-2000; 2000US-232081P.
XX 12-SEP-2000; 2000US-2321968P.
XX 14-SEP-2000; 2000US-232397P.
XX 14-SEP-2000; 2000US-232398P.
XX 14-SEP-2000; 2000US-232399P.
XX 14-SEP-2000; 2000US-232400P.
XX 14-SEP-2000; 2000US-232401P.
PR 14-SEP-2000; 2000US-233063P.
PR 14-SEP-2000; 2000US-233064P.
PR 14-SEP-2000; 2000US-233065P.
PR 21-SEP-2000; 2000US-234223P.
PR 21-SEP-2000; 2000US-234274P.
PR 25-SEP-2000; 2000US-234997P.
PR 25-SEP-2000; 2000US-234998P.
PR 26-SEP-2000; 2000US-235484P.
PR 27-SEP-2000; 2000US-235834P.
PR 27-SEP-2000; 2000US-235836P.
PR 29-SEP-2000; 2000US-236327P.
PR 29-SEP-2000; 2000US-236367P.
PR 29-SEP-2000; 2000US-236368P.
PR 29-SEP-2000; 2000US-236369P.
PR 29-SEP-2000; 2000US-236370P.
PR 02-OCT-2000; 2000US-236802P.
PR 02-OCT-2000; 2000US-237037P.
PR 02-OCT-2000; 2000US-237038P.
PR 02-OCT-2000; 2000US-237039P.
PR 13-OCT-2000; 2000US-239935P.
PR 13-OCT-2000; 2000US-239937P.
PR 20-OCT-2000; 2000US-240960P.
PR 20-OCT-2000; 2000US-241221P.
PR 20-OCT-2000; 2000US-241785P.
PR 20-OCT-2000; 2000US-241786P.
PR 20-OCT-2000; 2000US-241787P.
PR 20-OCT-2000; 2000US-241808P.
PR 20-OCT-2000; 2000US-241809P.
PR 20-OCT-2000; 2000US-241826P.
PR 01-NOV-2000; 2000US-244617P.
PR 08-NOV-2000; 2000US-246474P.
PR 08-NOV-2000; 2000US-246475P.
PR 08-NOV-2000; 2000US-246476P.
PR 08-NOV-2000; 2000US-246477P.
PR 08-NOV-2000; 2000US-246478P.
PR 08-NOV-2000; 2000US-246523P.
PR 08-NOV-2000; 2000US-246524P.
PR 08-NOV-2000; 2000US-246525P.
PR 08-NOV-2000; 2000US-246526P.
PR 08-NOV-2000; 2000US-246527P.
PR 08-NOV-2000; 2000US-246528P.
PR 08-NOV-2000; 2000US-246532P.
PR 08-NOV-2000; 2000US-246609P.
PR 08-NOV-2000; 2000US-246610P.
PR 08-NOV-2000; 2000US-246611P.
PR 08-NOV-2000; 2000US-246613P.
PR 17-NOV-2000; 2000US-249207P.
PR 17-NOV-2000; 2000US-249208P.
PR 17-NOV-2000; 2000US-249209P.
PR 17-NOV-2000; 2000US-249210P.
PR 17-NOV-2000; 2000US-249211P.
PR 17-NOV-2000; 2000US-249212P.
PR 17-NOV-2000; 2000US-249213P.
PR 17-NOV-2000; 2000US-249214P.
PR 17-NOV-2000; 2000US-249215P.
PR 17-NOV-2000; 2000US-249216P.
PR 17-NOV-2000; 2000US-249217P.
PR 17-NOV-2000; 2000US-249218P.
PR 17-NOV-2000; 2000US-249244P.
PR 17-NOV-2000; 2000US-249245P.
PR 17-NOV-2000; 2000US-249264P.
PR 17-NOV-2000; 2000US-249265P.
PR 17-NOV-2000; 2000US-249297P.
PR 17-NOV-2000; 2000US-249299P.
PR 17-NOV-2000; 2000US-249300P.
PR 01-DEC-2000; 2000US-250160P.
PR 01-DEC-2000; 2000US-250391P.
PR 05-DEC-2000; 2000US-251030P.
PR 05-DEC-2000; 2000US-251988P.
PR 05-DEC-2000; 2000US-256719P.
PR 06-DEC-2000; 2000US-251479P.
PR 06-DEC-2000; 2000US-251856P.

PR 08-DEC-2000; 2000US-251868P.
 PR 08-DEC-2000; 2000US-251869P.
 PR 08-DEC-2000; 2000US-251989P.
 PR 11-DEC-2000; 2000US-251990P.
 PR 11-DEC-2000; 2000US-254097P.
 PR 05-JAN-2001; 2001US-259678P.
 XX (HUMA-) HUMAN GENOME SCI INC.
 PA Rosen CA, Barash SC, Ruben SM;
 PI
 XX
 XX
 DR WPI: 2001-476161/51.
 DR P-PSDB; ABB10246.
 XX
 PT Isolated nucleic acid molecule encoding an inflammation-associated
 PT polypeptide is used in preventing, treating or ameliorating a medical
 PT condition -
 XX
 XX
 PS Claim 1; SEQ ID NO: 134; 859pp + Sequence Listing; English.
 XX
 CC The present invention provides human cDNAs, proteins and related genomic
 CC DNAs. These can be used in the treatment of neural, immune system,
 CC muscular, reproductive, gastrointestinal, pulmonary, cardiovascular,
 CC renal and proliferative disorders and inflammation. The present sequence
 CC is a cDNA of the invention.
 XX
 SO Sequence 5769 BP; 1299 A; 1695 C; 1628 G; 1146 T; 1 other;

alignment_scores: Quality: 9464.50 Length: 1762

Ratio: 5.427 Gaps: 1
 Percent Similarity: 98.978 Percent Identity: 98.978

alignment_block:

US-10-044-807-2 x ABA06468 ..

Align seg 1/1 to: ABA06468 from: 1 to: 5769

1 MetGluCysCysArgAlaThrProGlyThrLeuLeuLeuAl 17
 65 ATGGAATGCTGCGTGGGCAACTCTGGCACACTGCTCTTTCGGC 114
 17 aPheLeuLeuLeuSerSerArgThrAlaArgSerGluGluAspArgasp 34
 115 TTTCCTGCTCCTGAGTTCAGGACCGCACGCTCCGAGGAGGACCGGACG 164
 34 lyLeuTrpAspAlaTrpGlyProTyrSerGluCysSerArgThrCysgly 50
 165 GCCTATGGGATGCTGGGGCCCATGGAGTGAATGCTCAGCACCTGCGGG 214
 51 GlyClyAlaSerTyrSerLeuArgArgCysLeuSerSerLysSerCysGl 67
 215 GGTGGGCGCTCCTACTCTCTGAGCGCTGCTGAGCAGCAAGAGCTGTGA 264
 67 uGlyArgAsnIleArgTyrArgThrCysSerAsnValAspCysProProG 84
 265 AGAAGAAATATCCGATACAGAACATGCAGTAATGTGGACTGCCACACG 314
 84 luAlaGlyAspPheArgAlaGlnCysSerAlaHisAsnAspVallys 100
 315 AAGCAGGTGATTTCCGAGCTCAGCAATGCTCAGCTCATATGATGTCAG 364
 101 HisHisGlyGlnPheTyrGluTrpLeuProValSerAsnAspProAspAs 117
 365 CACCATGGCCAGTTTATGAATGGCTTCCCTGCTTAATGACCCGTGACAA 414
 117 nProCysSerLeuLysCysGlnAlaLysGlyThrThrLeuValGluGl 134
 415 CCCATGTTCACCTCAAGTGCCCAAGCAAGCAACCCCTGTTGTTGAAC 464
 134 euAlaProLysValLeuAspGlyThrArgCysTyrThrGluSerLeuAsp 150

465 TAGCACCTAAGGTCTTAGATGGTACGGCTTGCTATACAGAAATCTTTGGAT 514
 151 MetCysIleSerGlyLeuCysGlnIleValGlyCysAspHisGlnLeuGl 167
 515 ATGTGCATCAGTGTGTTATGCCAAATTTGCTGGCTCGGATCACCAGCTGG 564
 167 ySerThrValLysGluAspAsnCysGlyValCysAsnGlyAspGlyserT 184
 565 AAGCACCGTCAAGGAAGATAACTGTGGGTCTGCAACGGAGATGGGTCCA 614
 184 hrCysArgLeuValArgGlyGlnTyrLysSerGlnLeuSerAlaThrLys 200
 615 CCTGCCGCTGTGCGAGGGCAGTATAATCCAGCTCTCCGCAACCAAAA 664
 201 SerAspThrValValAlaIleProTyrGlySerArgHisIleArgLe 217
 665 TCGCATCATCTGTGGTTGCAATTCCTATGGAAGTAGACATATTCGCT 714
 217 uValLeuLysGlyProAspHisLeuTyrLeuGluThrLysThrLeuGlnG 234
 715 TGTCTTAAAGGTCCTGATCATCTTATCTGAAACCAAAACCCCTCCAGG 764
 234 lyThrLysGlyGluAsnSerLeuSerThrGlyThrPheLeuValAsp 250
 765 GGACTAAGGTGAACAGCTCAGCTCCACAGGAACCTTCTCTGTGGAC 814
 251 AsnSerSerValAspPheGlnLysPheProAspLysGluIleLeuArgMe 267
 815 AATTCTAGTGTGGACTTCCAGAAATTCACAGCAAAAGAGATACAGAAT 864
 267 tAlaGlyProLeuThrAlaAspPheIleValLysIleArgAsnSerGlyS 284
 865 GGCTGGACCACTCACAGCAGATTTCAATGTCAAGATTGTAACCTCGGCT 914
 284 erAlaAspSerThrValGlnPheIlePheTyrGlnProIleIleHisArg 300
 915 CCGCTGACAGTACAGTCCAGTTCATCTCTATCAACCATCATCCACGA 964
 301 TrpArgGlnThrAspPhePheProCysSerAlaThrCysGlyGlyGlyTY 317
 965 TGGAGGAGACGGATTTCTTCTTGTCTCAGCAACCTGTGGAGGAGGTTA 1014
 317 xGlnLeuThrSerAlaGluCysTyrAspLeuArgSerAsnArgValVala 334
 1015 TCAGCTGACATCGGCTGAGTCTACGATCTCAGGAGCAACCGTGTGGTGTG 1064
 334 laAspGlnTyrCysHisTyrTyrProGluAsnIleLysProLysProLys 350
 1065 CTGACCAATAGTGTCACTATTACCCAGAGAACATCAAAACCAACCCAG 1114
 351 LeuGlnGluCysAsnLeuAspProCysProAlaSerAspGlyTyrLysGl 367
 1115 CTTCAGGAGTGCAACTTGGATCTTGTCCAGCC..... 1147
 367 nIleMetProTyrAspLeuTyrHisProLeuProArgTrpGluAlaThrP 384
 1148AGTGGAGGCCACCC 1163
 384 roIrpThrAlaCysSerSerCysGlyGlyGlyIleGlnSerArgAla 400
 1164 CATGACCGCGTGTCTCTCTCTGTGGGGGGGCGCATCCAGAGCGGGCA 1213
 401 ValSerCysValGluGluAspIleGlnGlyHisValThrSerValGluGl 417
 1214 GTTTCCTGTGTGGAGGAGGACATCCAGGGGCACTGCTCAGTCCGAAGA 1263
 417 uTrpLysCysMetTyrThrProLysMetProIleAlaGlnProCysAsnI 434
 1264 GTGGAATGTCATGTACACCCCTAAGATGCCCATCCGAGCCCTCGCAACA 1313
 434 lePheAspCysProLysTrpLeuAlaGlnGluTrpSerProCysThrVal 450
 1314 TTTTGTACTGCCCTAAATGGCTGGCACAGGAGTGGTCTCCGTGCACAGTG 1363

451 ThrCysGlyGlnGlyLeuArgTyrArgValValLeuCysIleAspHisAr 467
 |||||
 1364 ACATGTGGCCAGGCGCTCAGATACCGTGTGGTCTCTGCATCGACCATCG 1413
 |||||
 467 gGlyMetHisThrGlyGlyCysSerProLysThrLysProHisIleLysG 484
 |||||
 1414 AGGAATGCACACAGGAGGCTGTAGCCCAAAACAAAGCCACATAAAG 1463
 |||||
 484 luGluCysIleValProThrProCysTyrLysProLysGluLysLeuPro 500
 |||||
 1464 AGNAATGCATCGTACCCACATCCCTGCTATAAACCCAAAGAGAACTTCCA 1513
 |||||
 501 ValGluAlaLysLeuProThrPheLysGlnAlaGlnGluLeuGluGln 517
 |||||
 1514 GTCGAGGCCAAGTGGCATGGTTCAAACAAGCTCAAGAGCTAGAAGAAGG 1563
 |||||
 517 YAlaAlaValSerGluGluProSerPheIleProGluAlaTrpSerAlaC 534
 |||||
 1564 AGCTGCTGTCTCAGAGAGGCGCTCGTTTCATCCCAAGAGCGCTGGTGGCCCT 1613
 |||||
 534 ySThrValThrCysGlyValGlyThrGlnValArgIleValArgCysGln 550
 |||||
 1614 GCACAGTCACCTGTGGTGTGGGACCCAGGTGCGAATAGTCAGTGGCAG 1663
 |||||
 551 ValLeuLeuSerPheSerGlnSerValAlaAspLeuProIleAspGluCy 567
 |||||
 1664 GTGCTCTGCTCTCTCTCAGTCCGTGGCTGACCTGCCTATTGACGAGTG 1713
 |||||
 567 sGluGlyProLysProAlaSerGlnArgAlaCysTyrAlaGlyProCysS 584
 |||||
 1714 TGAAGGCCCAAGCCAGCAGTCCCGAGCGCTGTTATGCAAGGCCCATGCA 1763
 |||||
 584 erGlyGluIleProGluPheAsnProAspGluThrAspGlyLeuPheGly 600
 |||||
 1764 GCGGGAAATTCCTGAGTCAACCCACAGACAGATGGCTCTTTGGT 1813
 |||||
 601 GlyLeuGlnAspPheAspGluLeuTyrAspTrpGluTyrGluGlyPheTh 617
 |||||
 1814 GGCCTGCAGGATTCGACGAGCTGTATGACTGGGAGTATAGGGGTTCCAC 1863
 |||||
 617 rLysCysSerCysSerCysGlyGlyValGlnGluAlaValValSerC 634
 |||||
 1864 CAAGTGTCCGAGTCCGTGGAGAGGTGTCAGAGGCTGTGTGAGCT 1913
 |||||
 634 ysLeuAsnLysGlnThrArgGluProAlaGluGluAsnLeuCysValThr 650
 |||||
 1914 GCTTGAACAACACAGACTCGGAGGCTGCTGAGGAGAACCTGTGCTGACC 1963
 |||||
 651 SerArgArgProProGlnLeuLeuLysSerCysAsnLeuAspProCysPr 667
 |||||
 1964 AGCGCGCGGCCCCACACAGCTCCTGAAGTCCCTGCAATTTGGATCCCTGCC 2013
 |||||
 667 oAlaArgTrpGluIleGlyLysTrpSerProCysSerLeuThrCysGlyV 684
 |||||
 2014 AGCAGGTGGGAAATGGCAAGTGGAGTCCATGTAGTCTACATGTGGGG 2063
 |||||
 684 alGlyLeuGlnThrArgAspValPheCysSerHisLeuLeuSerArgGlu 700
 |||||
 2064 TCGGCTTACAGACCAGAGAGCTCTTCTGCACCCACCTGCTTCCAGAGAG 2113
 |||||
 701 MetAsnGluThrValIleLeuAlaAspGluLeuCysArgGlnProLysPr 717
 |||||
 2114 ATGAATGAACAGTCACTCTGGCTGATGAGCTGTGTGCCAGGCCAAGCC 2163
 |||||
 717 oSerThrValGlnAlaCysAsnArgPheAsnCysProProAlaTrpTyrP 734
 |||||
 2164 CAGCAGGTGCAAGCTGTAAACCGCTTAAATTGCCCCCGACCGCTGGTACC 2213
 |||||
 734 roAlaGlnTrpGlnProCysSerArgThrCysGlyGlyValGlnLys 750
 |||||
 2214 CTGCACAGTGCACCGGTGTCCAGAACGTTGTGGCGGGGTGTTTCAGNAA 2263
 |||||

751 ArgGluValLeuCysLysGlnArgMetAlaAspGlySerPheLeuGluLe 767
 |||||
 2264 CGTGAGGTTCCTTTCAGACGCGCATGCTGATGSCAGCTTCTCTGGAGCT 2313
 |||||
 767 uProGluThrPheCysSerAlaSerLysProAlaCysGlnGlnAlaCysL 784
 |||||
 2314 TCCTTGAGACCTTCTGTTCAGTTCAAACCTCGCTGCCAGCAAGCATGCA 2363
 |||||
 784 ysLysAspAspCysProSerGluTrpLeuLeuSerAspTrpThrGluCys 800
 |||||
 2364 AGAAGATGACTGTCCAGCAGGTGGCTCTCTCAGACTGGACAGAGTGT 2413
 |||||
 801 SerThrSerCysGlyGluGlyThrGlnThrArgSerAlaIleCysArgLy 817
 |||||
 2414 TCCACAAGCTCGGGGAAGGACCCAGACTCGAAGCGCCATTTGCCGAAA 2463
 |||||
 817 sMetLeuLysThrGlyLeuSerThrValValAsnSerThrLeuCysProp 834
 |||||
 2464 GATGCTGAAAACCGGCCCTCTCAACGGTTGTCAATTCACCCCTGTGCCGC 2513
 |||||
 834 roLeuProPheSerSerIleArgProCysMetLeuAlaThrCysAla 850
 |||||
 2514 CCCTGCCCTTCTCTCTCCATCAGGCCCTGTATGCTGGCAACCTGTGCA 2563
 |||||
 851 ArgProGlyArgProSerThrLysHisSerProHisIleAlaAlaAlaAr 867
 |||||
 2564 AGGCCCGGGCGCCATCCAGGAAGCACAGCCCGCACATCGCGGCCGCCAG 2613
 |||||
 867 gLysValIleGlnThrArgArgGlnArgLysLeuHisPheValValG 884
 |||||
 2614 GAAGGCTCTACATACAGACTCGCAGGCAGAGGAAGCTGCACCTTCGTGGTG 2663
 |||||
 884 lyGlyPheAlaTyrLeuLeuProLysThrAlaValValLeuArgCysPro 900
 |||||
 2664 GGGCTTCGCCCTACCTGCTCCCAAGACGGCGGTGGTCTGGCTGGCCGG 2713
 |||||
 901 AlaArgArgValArgLysProLeuIleThrTrpGluLysAspGlyGlnHi 917
 |||||
 2714 GCGGCGAGGGTCCGCAAGCCCTCATCACCTGGGAGAAGGACGGCCAGCA 2763
 |||||
 917 sLeuIleSerSerThrHisValThrValAlaProPheGlyTyrLeuLysI 934
 |||||
 2764 CCTCATCAGCTCGACGCGCACGTACGGTGGCCCCCTTCGGCTATCTCAAGA 2813
 |||||
 934 leHisArgLeuLysProSerAspAlaGlyValTyrThrCysSerAlaGly 950
 |||||
 2814 TCCACCGGCTCAAGCCCTCGGATCGAGCGCTCTACACCTGCTCAGCGGC 2863
 |||||
 951 ProAlaArgGluHisPheValIleLysLeuIleGlyGlyAsnArgLysLe 967
 |||||
 2864 CCGGCCCGGGAGACATTTGTATTAAAGTCTATCGAGGCAACCGCAAGCT 2913
 |||||
 967 uValAlaArgProLeuSerProArgSerGluGluGluValLeuAlaGlyA 984
 |||||
 2914 CGTGCCCGCGCCCTTGAGCCCGCAAGCTGAGAAAGAGTGTCTCGGGGGA 2963
 |||||
 984 rgLysGlyGlyProLysGluAlaLeuGlnThrHisLysHisGlnAsnGly 1000
 |||||
 2964 GGAAGGGGGCCCAAGAGGCCCTTCAGACCCCAACACACAGAACGGGG 3013
 |||||
 1001 IlePheSerAsnGlySerLysAlaGluLysArgGlyLeuAlaAlaAsnPr 1017
 |||||
 3014 ATCTTCTCAAACGCGCAGCAAGCGGAGAAAGCGGGCCCTGGCCGCCAACCC 3063
 |||||
 1017 oGlySerArgTyrAspAspLeuValSerArgLeuLeuGlnGlyGlyT 1034
 |||||
 3064 GGGAGCGCGCTACGACGACCTCGTCTCCCGGCTGCTGGAGCAGGGCGGCT 3113
 |||||
 1034 rpProGlyGluLeuAlaSerTrpGluAlaGlnAspSerAlaGluArg 1050
 |||||
 3114 GGCCCGGAGAGCTGCTGGCCCTCGTGGGAGGCGCAGGACTCTGCGGAAAGG 3163
 |||||
 1051 AsnThrThrSerGluGluAspProGlyAlaGluGlnValLeuLeuHisIle 1067
 |||||

|||||
3164 AACACGACCTCGGAGAGGACCCGGGTGCAGAGCAAGTGCTCCCTGCACCT 3213
1067 uProPheThrMetValThrGluGlnArgArgLeuAspAspIleLeuGlyA 1084
|||||
3214 GCCCTTACCATGCTGACCGAGCAGCGCGCTGCAGACATCTCGGGA 3263
1084 snLeuSerGlnGlnProGluGluLeuArgAspLeuTyrSerLysHisLeu 1100
3264 ACCTCTCCACGACCGCGAGAGCTGCGCGACCTCTACAGCAAGCACCTG 3313
1101 ValAlaGlnLeuAlaGlnGluIlePheArgSerHisLeuGluHisGlnAs 1117
|||||
3314 GTGCCCGAGCTGGCCCGAGGAGATCTCCGACGACCCTGGAGCACCAAG 3363
1117 pThrLeuLeuLysProSerGluArgArgThrSerProValThrLeuSerP 1134
3364 CACGCTCCTGAAGCCCTCGGAGCGCAGGACTTCCCGAGTGACTCTCTCGC 3413
1134 roHisLysHisValSerGlyPheSerSerSerLeuArgThrSerSerThr 1150
3414 CTCAAAACACGTGCTGGGTTTCAGCAGCTCCCTGCGGACCTCTCCACC 3463
1151 GlyAspAlaGlyGlySerArgArgProHisArgLysProThrIleLe 1167
3464 GGGACCGCGGGGAGGCTCTCGAAGGCCACACCGCAAGCCCACTCCT 3513
1167 uArgLysIleSerAlaAlaGlnGlnLeuSerAlaSerGluValValThrH 1184
3514 GCGCAAGATCTACGCGCCCGACGAGCTCTCAGCCTCGGAGTGTCAACC 3563
1184 lsLeuGlyGlnThrValAlaLeuAlaSerGlyThrLeuSerValLeuLeu 1200
3564 ACCTGGGGCAGACGGTGGCCCTGGCCAGCGGACACTCAGTGTTCCTCTG 3613
1201 HisCysGluAlaIleGlyHisProArgProThrIleSerIrpAlaArgAs 1217
3614 CACTGTGAGGCCATCGCCACCCACCCAGGCCCTACCATCAGCTGGGCCAGGA 3663
1217 nGlyGluGluValGlnPheSerAspArgIleLeuLeuGlnProAspAsps 1234
3664 TGGAGAAGAAGTTCAGTTCAGTGACAGGATTCTTCTACAGCAGATGATT 3713
1234 erLeuGlnIleLeuAlaProValGluAlaaspValGlyPheTyrThrCys 1250
3714 CCTTACAGATCTTGGCACCACTAGTGAACAGATGTGGGTTTCTACACTGC 3763
1251 AsnAlaThrAsnAlaLeuGlyTyrAspSerValSerIleAlaValThrLe 1267
3764 AATGCCACCAATGCTTGGGATAGGACTCTGTCTCCATTGCCGTACATT 3813
1267 uAlaGlyLysProLeuValLysThrSerArgMetThrValIleAsnThrG 1284
3814 AGCAGGAAGCCACTAGTGAAGAGTCCACGAATGACAGTGATCAACACGG 3863
1284 LuLysProAlaValThrValAspIleGlySerThrIleLysThrValGln 1300
3864 AGAAGCTGGAGTCCAGTCCATATAGGAAGCACCATCAAAACAGTGCAG 3913
1301 GlyValAsnValThrIleAsnCysGlnValAlaGlyValProGluAlaG 1317
3914 GGAGTGAATGTGACAATCACTGCCAGGTTGCAGGAGTGCCTGAAGCTGA 3963
1317 uValThrTrpPheArgAsnLysSerLysLeuGlySerProHisHisLeuH 1334
3964 AGTCACCTGGTTCAGGAATAAAAGCAAACTGGGCTCCCGCAGCACCATCTGC 4013
1334 isGluGlySerLeuLeuLeuThrAsnValSerSerSerAspGlnGlyLeu 1350
4014 ACGAAGGCTCCTGCTCCTCACAACAGTGCTCCTCGGATCAGGGCCTG 4063
1351 TyrSerCysArgAlaAlaAsnLeuHisGlyGluLeuThrGluSerThrGl 1367
|||||

4064 TACTCCTCGACGGGGGCCAAATCTTCAATGGAGAGCTGACTGAGACACCCA 4113
1367 nLeuLeuIleLeuAspProProGlnValProThrGlnLeuGluAspIleA 1384
|||||
4114 GCTGCTGATCTAGATCCCCCCAAGTCCCCACACAGTTGGAAGACATCA 4163
1384 rgAlaLeuLeuAlaAlaThrGlyProAsnLeuProSerValLeuThrSer 1400
4164 GGGCTTGTCTGCTGCCACTGGACCGAACTTCCCTCAGTGTGACGTCT 4213
1401 ProLeuGlyThrGlnLeuValLeuAspProGlyAsnSerAlaLeuLeuGl 1417
4214 CCTCTGGGAACACAGCTGGTCTGGATCCTGGGAATCTGCTCTCCTTGG 4263
1417 yCysProIleLysGlyHisProValProAsnIleThrTrpPheHisGlyG 1434
4264 CTGCCCCATCAAGGTCAACCTGTCCCTTAATATCACTGGTTTCATGGTG 4313
1434 lyGlnProIleValThrAlaThrGlyLeuThrHisHisIleLeuAlaAla 1450
4314 GTCAGCCAATTGTCACCTGCCACAGGACTGACGATCACATCTTGGCAGCT 4363
1451 GlyGlnIleLeuGlnValAlaAsnLeuSerGlyGlySerGlnGlyGluPh 1467
4364 GGACAGATCCTTCAAGTTGCAACCTTAGCGGTGGGTCTCAAGGGGAATT 4413
1467 eSerCysLeuAlaGlnAsnGluAlaGlyValLeuMetGlnLysAlaSerL 1484
4414 CAGCTGCCCTTGTCTCAGAAATGAGGAGGGGTGCTCATGCAAGAGGCTATT 4463
1484 euValIleGlnAspTyrTrpSerValAspArgLeuAlaThrCysSer 1500
4464 TAGTGATCCAAGATTACTGGTGGTCTGTGGACAGACTGGCAACCTGTCTCA 4513
1501 AlaSerCysGlyAsnArgGlyValGlnGlnProArgLeuArgCysLeuLe 1517
4514 GCCCTCTGTGTAACTGGGGGGTTTCAGCAGCCCGCTTGGAGTGGCTGCT 4563
1517 uAsnSerThrGluValAsnProAlaHisCysAlaGlyLysValArgProA 1534
4564 GAACAGCAGCGAGGTCAACCTTGCCACTGCGCAGGGAAGGTTCGCCCTG 4613
1534 laValGlnProIleAlaCysAsnArgArgAspCysProSerArgTrpMet 1550
4614 CGGTGCAGCCCATCGGTGCAACCGGAGAGACTGCCCTTCTCGGTGGATG 4663
1551 ValThrSerTrpSerAlaCysThrArgSerCysGlyGlyValGlnTh 1567
4664 GTGACCTCTGCTGCTGCTGTACCCGGAGCTGTGGGGAGGTGTCCAGAC 4713
1567 rArgArgValThrCysGlnLysLeuLysAlaSerGlyIleSerThrProv 1584
4714 CCGCAGGGTGACCTGTCAAAAGCTGAAAGCTCTGGGATCTCCACCCCTG 4763
1584 alSerAsnAspMetCysThrGlnValAlaLysArgProValAspThrGln 1600
4764 TGTCCAATGACATGTGACCCAGGTGCGCAAGCGGCTGTGGACACCCAG 4813
1601 AlaCysAsnGlnGlnLeuCysValGluIrpAlaPheSerSerTrpGlyGl 1617
4814 GCCTGTAAACAGCAGCTGTGTGTGAGGTGGCTTCTCCAGCTGGGGCA 4863
1617 nCysAsnGlyProCysIleGlyProHisLeuAlaValGlnHisArgGlnV 1634
4864 GTGCAATGGGCTTGCATCGGGCTCACTAGCTGTGCAACACAGACAAG 4913
1634 alPheCysGlnThrArgAspGlyIleThrLeuProSerGluIleCysSer 1650
4914 TCTTCTGCCACACAGCGGATGGCATCCTTACCATCAGAGCAGTGCAGT 4963
1651 AlaLeuProArgProValSerThrGlnAsnCysTrpSerGluAlaCysSe 1667
4964 GCTCTTCGACAGCCTGTGAGCACCCAGCAACTGCTGTCAGAGGCTGCAG 5013

1667 rValHisTrpArgValSerLeuTrpThrLeuCysThrAlaThrCysGlyA 1684
 |||||
 5014 TGTACACTGGAGAGTCAGCTGTGGACCTGTGCACAGCTACCTGTGGCA 5063
 1684 sNtYrGlyPheGlnSerArgValGluCysValHisAlaArgThrAsn 1700
 |||||
 5064 ACTACGGCTTCAGTCCCGCGGTGGAGTGTGTGCATGCCCGCACCAAC 5113
 1701 LysAlaValProGluHisLeuCysSerTrpGlyProArgProAlaAsnTr 1717
 |||||
 5114 AAGGCAGTGCCTGAGCACCTGTCTCTGGGGCCCGCGCTGCCACTG 5163
 1717 pGlnArgCysAsnIleThrProCysGluAsnMetGluCysArgAspThrT 1734
 |||||
 5164 GCAGCGCTGCACATCACCCCATGTGAAACATGGAGTGCAGACACCA 5213
 1734 hrArgTrpCysGluLysValLysGlnLeuLysCysGlnLeuSerGln 1750
 |||||
 5214 CCAGGTACTGCCGAGAGGTGAACAGCTGAAACTGTGCCAAGCTCAGCCAG 5263
 1751 PheLysSerArgCysCysGlyThrCysGlyLysAla 1762
 |||||
 5264 TTTAAATCTCGTGTGCTGGAAGTGTGGCAAGCG 5299

seq_name: /SIDS1/gcgdata/hold-geneseq/geneseqn-emb1/NA2001A.DAT:AAF97926

seq_documentation_block:

ID AAF97926 standard; cDNA: 3559 BP.

XX AC AAF97926;

XX DT 01-JUN-2001 (first entry)

XX DE Human secreted protein cDNA, SEQ ID NO: 53.

XX KW Human; secreted protein; immunomodulatory; antisclerotic;
 KW dermatological; antiinflammatory; anti-HIV; cytostatic; cardiant;
 KW vascular; anti-angiogenic; ophthalmological; neuroprotectant;
 KW neotropic; anticonvulsant; anti-Alzheimer's; antiparkinsonian;
 KW antimicrobial; vulnerary; vaccine; gene therapy; cancer;
 KW protein coordinate data; infection; ss.

XX OS Homo sapiens.

XX FN WO200121658-A1.

XX PD 29-MAR-2001.

XX PF 22-SEP-2000; 2000WO-US26013.

XX PR 24-SEP-1999; 99US-0155709.

XX PA (HUMA-) HUMAN GENOME SCI INC.

XX PI Ni J, Baker KP, Birse CE, Ebner R, Fiscella M, Komatsoulis GA;
 PI Lafleur DW, Moore PA, Olsen HS, Rosen CA, Ruben SA, Soppet DR;
 PI Young PE, Wei P, Florence KA;

XX DR WPI; 2001-235311/24.

XX PT Nucleic acids encoding 32 human secreted polypeptides, useful for
 PT preventing, diagnosing and/or treating e.g. cancers, Parkinson's
 PT disease and diabetic retinopathy -

XX PS Claim 1; Page 747-748; 890pp; English.

XX CC The present sequence encodes one of 32 novel human secreted polypeptides.
 CC The nucleic acid molecules and polypeptides they encode may be used in
 CC the prevention, diagnosis and treatment of diseases such as
 CC immune disorders (e.g. multiple sclerosis, systemic lupus erythematosus
 CC and human immuno-deficiency virus (HIV) infections), hyperproliferative
 CC disorders (e.g. cancers and Gaucher's disease), cardiovascular diseases

CC (e.g. Scimitar syndrome, Chaga's cardiomyopathy and coronary
 CC arteriosclerosis), angiogenic disorders (e.g. corneal graft
 CC neovascularisation and diabetic retinopathy), neurological disorders
 CC (e.g. Huntington's chorea, Alzheimer's disease and Parkinson's disease),
 CC infectious diseases and/or for promoting wound healing, regeneration
 CC and/or chemotaxis. The nucleic acid molecules may be used to produce the
 CC secreted polypeptides. They may also be used as DNA probes in diagnostic
 CC assays to detect and quantitate the presence of similar nucleic acid
 CC sequences in samples. The polypeptides may be used as antigens in the
 CC production of antibodies and in assays to identify modulators of
 CC their expression and activity.

XX SQ Sequence 3559 BP; 807 A; 1037 C; 1005 G; 698 T; 12 other;

alignment_scores:

Quality: 6176.50 Length: 1201

Ratio: 5.279 Gaps: 3

Percent Similarity: 97.419 Percent Identity: 96.919

alignment_block:

US-10-044-807-2 x AAF97926 ..

Align seg 1/1 to: AAF97926 from: 1 to: 3559

73 TyrArgThrCysSerAsnValAspCysProGluAlaGlyAspPheAr 89
 :: ||||| ::::: |||||
 13 TTTTAAACATGCAGCCCTTTGGAAATGCCACCAAAAGCMSGTGTATTCG 62
 89 gAlaGlnGlnCysSerAlaHisAsnValLysHisGlyGlnPheT 106
 |||||
 63 AGCTCAGCAATGCTCAGCTCATATGTCAGCACCAGCCAGTGT 112
 106 yrGluTrpLeuProValSerAsnAspProAspAsnProCysSerLeuLys 122
 |||||
 113 ATGAATGGCTTCTGTGTCTTAATGACCCCTGACCAACCCATGTTCACTCAAG 162
 123 CysGlnAlaLysGlyThrLeuValGluLeuAlaProLysValle 139
 |||||
 163 TGCCAAAGCCAAAGAACACCCCTGTTGTGTGTAAGTACAGCCTAAGTCTT 212
 139 uAspGlyThrArgCysTrpThrGluSerLeuAspMetCysIleSerGlyL 156
 213 AGATGTGACGGTGTGCTATACAGAAATCTTTGGATATCTGCATCAGTGT 262
 156 euCysGlnIleValGlyCysAspHisGlnLeuGlySerThrValLysGlu 172
 |||||
 263 TATGCCAAATGTTGGCTGGATCACCAGCTGGGAAGCACCCTCAAGGAA 312
 173 AspAsnCysGlyValCysAsnGlyAspGlySerThrCysArgLeuValAr 189
 313 GATAACTGTGGGTCTGCAACGGAGATGGTCCACCTGCCGGCTGGTCCG 362
 189 gGlyGlnTrpLysSerGlnLeuSerAlaThrLysSerAspThrValV 206
 363 AGGGCAGTATAATCCAGCTCTCCGCAACCAATCGGATGATGACTGTG 412
 206 aAlaIleProTrpGlySerArgHisIleArgLeuValLysGlyPro 222
 413 TTGCAATTCCTATGGAAGTAGACATATTCGCCCTTGTCTTAAAGGTCCT 462
 223 AspHisLeuTrpLeuGluThrLysThrLeuGlnGlyThrLysGlyGluAs 239
 463 GATCACTTATATCTGGAACCAAAACCCCTCCAGGGGACTAAAGGTCAAAA 512
 239 nSerLeuSerSerThrGlyThrPheLeuValAspAsnSerSerValAsp 256
 513 CAGTCTCAGCTCCACAGGAACCTTCTCTGGACAATTTCTAGTGTGACT 562
 256 heGlnLysPheProAspLysGluIleuArgMetAlaGlyProLeuThr 272
 563 TCCAGAAATTTCCAGAACAAAGAGATACTAGAAATGGTGGACCACTACA 612

273 AlaAspPheIleValIysIleArgAsnSerGlySerAlaAspSerThrVa 289
|||||GCAGATTTCAATTCAGATTGTAACCTCGGGCTCCGCTCACAGTACAGT 662
289 lGlnPheIlePheTyrGlnProIleIleHisArgTrpArgLutThrAspP 306
|||||CCAGTTTCATCTTCAATCAACCCCATCATCCACCCATGAGGAGAGCGATT 712
306 hePheProCysSerAlaThrCysGlyGlyTyrGlnLeuThrSerAla 322
|||||TCWTTCCTTGCTCAGCAACTGTGGAGAGGTTATCAGTCGACATCGGCT 762
323 GluCysTyrAspLeuArgSerAsnArgValValAlaAspGlnTyrCysH 339
|||||GAGTGTACGATCTGAGGAGCAACCGTGTGTGCTGACCAATACTGTCA 812
339 sTyrTyrProGluAsnIleLysProLysProLysLeuGlnGluCysAsnL 356
|||||CTATTACCCAGAGAAATCAAAACCCCAACCCAAAGCTTCAGGAGTGCAACT 862
356 euAspProCysProAlaSerAspGlyTyrIysGlnIleMetProTyrAsp 372
|||||TGGATCTTGTCCAGCC..... 879
373 LeuTyrHisProLeuProArgTrpGluAlaThrProTrpThrAlaCysSe 389
|||||.....AGTGGGAGGCCACCCCATGGACCGCGTGTCTC 911
389 rSerSerCysGlyGlyIleGlnSerArgAlaValSerCysValGluG 406
|||||CTCCTCGTGGGGGGGGATCCAGAGCGGGCAGTTTCCGTGTGGAGG 961
406 luAspIleGlnGlyHisValThrSerValGluGluTrpLysCysMetTyr 422
|||||AGGACATCCAGGGGCATGTCACCTCAGTGGAGAGTGGAAATGCATGTAC 1011
423 ThrProLysMetProIleAlaGlnProCysAsnIlePheAspCysProLy 439
|||||ACCCTTAAGATGCCCATCGCGCAGCCCTGCACATTTTGTACTGCGCTAA 1061
439 sTrpLeuAlaGlnGluTrpSerProCysThrValThrCysGlyGlnGlyL 456
|||||ATGGCTGGACACAGAGTGTCTCCGTGCACAGTGACATGGCGCAGGGCC 1111
456 euArgTyrArgValValLeuCysIleAspHisArgGlyMetHisThrGly 472
|||||TCAGATACCGTGTGTCTCTGATCGACCATCGAGGAATGCACACAGGA 1161
473 GlyCysSerProLysThrLysProHisIleLysGluGluCysIleValPr 489
|||||GGCTGTAGCCCCAAAACAAAGCCCCACATAAAAGAGGAATGTCATCGTACC 1211
489 oThrProCysTyrLysProLysGluLysLeuProValGluAlaLysLeup 506
|||||CATTCTCTGCTATPAACCCAAAGAGAAACTTCCAGTCGAGGCCAAGTTGC 1261
506 roTrpPheLysGlnAlaGlnGluLeuGluGluGlyAlaAlaValSerGlu 522
|||||CATGCTTCAAAACAAGCTCAAGAGCTAGAGAAGAGAGCTGCTGTGTCAGAG 1311
523 GluProSerPheIleProGluAlaTrpSerAlaCysThrValThrCysGl 539
|||||GAGCCCTGCTTCAATTCARAGGCTGTGGCCCTGCACAGTCACTACCTGTGG 1361
539 yValGlyThrGlnValArgIleValArgCysGlnValLeuLeuSerPheS 556
|||||TGTGGGGACCCARCTCGGAATAGTCAAGTGCCAGGTGCTCTCTCTTCT 1411
556 erGlnSerValAlaAspLeuProIleAspGluCysGluGlyProLysPro 572
|||||CTCAGTCCCGTGGTGAACCTGCTATTGACGAGTGTGAAGGGCCCAAGCCA 1461
573 AlaSerGlnArgAlaCysTyrAlaGlyProCysSerGlyGluIleProGl 589

1462 GCATCCACAGCGTGCTGTATTATGAGGCCCATGTCAGCGGGGAAATTCCTGA 1511
589 uPheAsnProAspGluThrAspGlyLeuPheGlyGlyLeuGlnAspPheA 606
|||||GTTCAACCCACAGACAGACAGATGGGCTTTTGGTGGCTGCAGGATTCG 1561
606 spGluLeuTyrAspTrpGluTyrGluGlyPheThrLysCysSerGluSer 622
|||||ACGAGCTGTATGACTGGAGTATGAGGGGTTTCCCAAGTGTCTCGAGTCC 1611
623 CysGlyGlyGlyValGlnGluAlaValIserCysLeuAsnIysGlnTh 639
|||||TGTGGAGAGGTGTCCAGGAGGCTGTGCTGAGCTGCTTGAACAAACAGAC 1661
639 rArgGluProAlaGluAsnLeuCysValThrSerArgArgProProG 656
|||||TCGGGAGCCCTGCTGAGAGAACCTGTGCTGACACGCGCGCGGCCCCAC 1711
656 lNLeuLeuLysSerCysAsnLeuAspProCysProAlaArgTrpGluIle 672
|||||AGCTCCTGAAGTCTGCAATTTGGATCCCTGCCAGCAAGGTGGGAATT 1761
673 GlyLysTrpSerProCysSerLeuThrCysGlyValGlyLeuGlnThrAr 689
|||||GGCAAGTGGAGTCCATAGTCTCAGTGTGGGTGCGCTTACAGACAG 1811
689 qAspValPheCysSerHisLeuLeuSerArgGluMetAsnGluThrValI 706
|||||AGAGCTTCTTGCAGCACCTGCTTCCAGAGAGATGAATGAACAGTCA 1861
706 leLeuAlaAspGluLeuCysArgGlnProLysProSerThrValGlnAla 722
|||||TCCTGGCTGATGAGCTGTGTGCCAGCCCAAGCCAGCACGGTGCAGCT 1911
723 CysAsnArgPheAsnCysProProAlaTrpTyrProAlaGlnTrpGlnPr 739
|||||TGTAAACCGCTTAAATTTGCCCGCCCGGCTGTACCTGCACAGTGGCAGCC 1961
739 oCysSerArgThrCysGlyGlyValGlnLysArgGluValLeuCysL 756
|||||GTGTTCCAGAACGTTGTGGGGGGTGTTCAGAAACGTTGAGGTTCTTGCA 2011
756 ySglnArgMetAlaAspGlySerPheLeuGluLeuProGluThrPheCys 772
|||||AGCAGCGCATGGCTGATGGCAGCTTCTGGAGCTTCTGAGACCTTCTGT 2061
773 SerAlaSerLysProAlaCysGlnAlaCysLysLysAspCysPr 789
|||||TCAGCTTCAAAACCTGCTGCCAGCAAGCATGCAAGAAAGATGACTGTCC 2111
789 oSerGluTrpLeuLeuSerAspTrpThrGluCysSerThrSerCysGlyG 806
|||||CAGCGAGTGGCTTCTCTCAGACTGGCAGAGTGTTCACCAAGCTGCGGGG 2161
806 luGlyThrGlnThrArgSerAlaIleCysArgLysMetLeuLysThrGly 822
|||||AAGSCACCCAGACTCGAAGCGCCATTTGCCAAGAGATGCTGAACACCGGC 2211
823 LeuSerThrValValAsnSerThrLeuCysProProLeuProPheSerSe 839
|||||CTCTCAACGGTTGTCAATTCACCTGTGCGCGCCCTTCTCTCTTC 2261
839 rSerIleArgProCysMetLeuAlaThrCysAlaArgProGlyArgProS 856
2262 CTCCATCAGCCCTGTATGTGTGGCAACTGTGCAAGCCCGCGCGCCCAT 2311
856 erThrLysHisSerProHisIleAlaAlaAlaArgLysValTyrIleGln 872
2312 CCAGGAAGCAGACCGCCACATCGCGCGCCGAGGAAGGTCTACATACAG 2361
873 ThrArgArgGlnArgLysLeuHisPheValValGlyGlyPheAlaTyrLe 889
|||||

2362 ACTCGCAGGAGAGAAAGCTGCCTCGTGGTGGGGGCTTCGCCTACCT 2411
889 uLeuProLysThrAlaValValLeuArgCysProAlaArgValArgL 906
|||||
2412 GCTCCCAAGAGCGCGTGTGTGCGCTGCCGCGCGAGGTCGCA 2461
906 ysProLeuIleThrTrpGluLysAspGlyGlnHisLeuIleSerSerThr 922
|||||
2462 AGCCCTCATCAGCTGGGAGAGGAGCGCCGAGCAGCTCATCAGCTCGACG 2511
923 HisValThrValAlaProPheGlyTyrLeuLysTleHisArgLeuLysPr 939
|||||
2512 CAGTCAGGTGGCCCTTCGCGTATCTCAAGATCCACCGCTCAAGCC 2561
939 oSerAspAlaGlyValTyrThrCysSerAlaGlyProAlaArgGluHisP 956
|||||
2562 CTCGGATGCAGCGCTCTACACCTGCTCAGCGGGCCGCGGAGCACT 2611
956 heValIleLysLeuIleGlyClyAsnArgLysLeuValAlaArgProLeu 972
|||||
2612 TTGTGATTAAAGCTCATCGGAGGCAACCGCAAGCTCGTGGCCGCGCCCTTG 2661
973 SerProArgSerGluGluGluValLeuAlaGlyValArgLysGlyProLy 989
|||||
2662 AGCCCGAAGACTCAGGAAGAGTGTCTGCGGGAGGAGGCGGCCGAA 2711
989 sGluAlaLeuGlnThrHisLysHisGlnAsnGlyIlePheSerAsnGlyS 1006
|||||
2712 GGAGGCCCTGCAGACCCACAAACACCAAGACGGGATCTTCTCCAACGGCA 2761
1006 erLysAlaGluLysArgGlyLeuAlaAlaAsnProGlySerArgTyrAsp 1022
|||||
2762 GCAAGCGGAGAGAGCGGGCTGCGCCCAACCCGGGAGCGCTACGAC 2811
1023 AsPleuValSerArgLeuLeuGluGlnGlyTyrProGlyGluLeuLe 1039
|||||
2812 GACCTGCTCTCCGCTGCTGGAGCAGGCGGCTGGCCCGGAGAGCTGCT 2861
1039 uAlaSerTrpGluAlaGlnAspSerAlaGluArgAsnThrThrsrGluG 1056
|||||
2862 GGCCTCTGGGAGCGCAGGACTCTGCGGAAAGGAACACGACCTCGGAGG 2911
1056 luAspProGlyAlaGluGlnValLeuLeuHisLeuProPheThrMetVal 1072
|||||
2912 AGGACCGGCTGAGAGCAAGTGTCTGCACTGCGCTTCCACCATGGTG 2961
1073 ThrGluGlnArgArgLeuAspAspIleLeuGlyAsnLeuSerGlnGlnPr 1089
|||||
2962 ACCGAGCGCGGCGCTGGAGCAGCATCTCTGGGNAACCTCTCCACGACGCC 3011
1089 oGluGluLeuArgAspLeuTyrSerLysHisLeuValAlaGlnLeuAlaG 1106
3012 CGAGGAGCTGCGGCGACCTCTACAGCAAGCACCCTGGTGGCCAGCTGGGCC 3061
1106 lnGluIlePheArgSerHisLeuGluHisGlnAspThrLeuLeuLysPro 1122
|||||
3062 AGGAGATCTTCCGAGCCACTGGAGCACCAGGACAGCTGCTCGTAAGCCC 3111
1123 SerGluArgArgThrSerProValThrLeuSerProHisLysHisValse 1139
|||||
3112 TCGGAGCGCAGGACTTCCCGAGTGACTCTCTCGCTCATTAACACAGCTGTC 3161
1139 rGlyPheSerSerSerLeuArgThrSerSerThrGlyAspAlaGlyGlyG 1156
3162 TGGCTTTCAGCAGCTCCCTGGGAGCTCTCTCACCGGGGAGCGCGGGGAG 3211
1156 lySerArgArgProHisArgLysProThrIleLeuArgLysIleSerAla 1172
3212 GCTCTGAAGGCCACACCGAAGCCCACTCTCTGCGCAAGATCTCAGCG 3261
1173 AlaGlnGlnLeuSerAlaSerGluValValThrHisLeuGlyGlnThrVa 1189
|||||
3262 GCCCAGCAGCTCTCAGCCTCGGAGTGGTCACCCACCTGGGGCAGACGGT 3311

1189 lAlaLeuAlaSerGlyThrLeuSerValLeuLeuHisCysGluAlaIleG 1206
|||||
3312 GGCCCTGGCCAGCGGACACTGAGTGT.CTTCTGCACTGTGAGGCCATCG 3360
1206 lyHisProArgProThrIleSerTrpAlaArgAsnGlyGluGluValGln 1222
|||||
3361 GCNACCCAAAGGCCCTACCATCAGCTGGCGCAGGAATGGAGAAGAAGT.CAG 3409
1223 PheSerAspArgIleLeuLeuGlnProAspAspSerLeuGlnIleLeu.A 1239
|||||
3410 TTCAGTGACAGGATCTCTACAGSCAGATGATTCYTTACAGATCTTGGG 3459
1239 laProValGlu.AlaAspValGlyPheTyrThrCysAsnAlaThrAsnAl 1255
|||||
3460 CACCAGTGGAAAGCAGATGGGTTTCTACACTTGCATGTCAC.AATGC 3508
1255 aLeuGlyTyrAspSerValSerIleAlaValThrLeuAlaGlyLysPro 1271
|||||
3509 CTTGGGATACGAMTTTGTCTCCATTGGCGTCACATTACAGAAAGCCC 3557

seq_name: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA2000.DAT.AAA47482

seq_documentation_block:

ID AAA47482 standard; cDNA; 2689 bp.

XX AAA47482;

XX AC

XX DT 20-OCT-2000 (first entry)

XX XX Human TANGO 224 coding sequence (form 1).

KW TANGO; 128; 140; 197; 212; 213; 224; 239; modulating agent; asthma;
KW graft versus-host diseases; rheumatoid arthritis; psoriasis;
KW inflammatory bowel disease; septic shock; ulcerative colitis;
KW Crohn's disease; chronic myelogenous leukemia; cancer; liver
KW disease; Hodgkin's disease; osteoarthritis; Lyme's disease;
KW cachexia; autoimmune disease; myasthenia gravis; autoimmune diabetes;
KW systemic lupus erythematosus; transgenic animal; diagnosis;
KW prognosis; prophylactic; therapeutic; human; ds.

XX OS Homo sapiens.

XX FH

XX Key Location/Qualifiers

XX CDS 67..2688

XX FT /*tag= a

XX FT /product= TANGO 224

XX PN WO200039284-A1.

XX PD 06-JUL-2000.

XX PF 23-DEC-1999; 99WO-US31025.

XX PR 30-DEC-1998; 98US-0223546.

XX XX (MILL-) MILLENNIUM PHARM INC.

XX XX Holtzman DA;

XX XX WPI; 2000-465743/40.

XX DR P-PSDB; AAB01431.

XX DR Novel nucleic acid sequences encoding TANGO-128, 140, 197, 212, 213,
XX PT 224 and 239 polypeptides useful for the treatment of asthma, rheumatoid
XX PT arthritis, psoriasis and autoimmune diseases

XX PS Claim 1; Fig 30; 209pp; English.

XX CC Nucleic acids encoding TANGO polypeptides are useful as modulating
XX CC agents for regulating cellular processes like asthma, graft
XX CC versus-host diseases, rheumatoid arthritis, psoriasis, inflammatory
XX CC bowel disease, septic shock, ulcerative colitis, Crohn's disease,

CC chronic myelogenous leukemia, cancer, liver disease, Hodgkin's
 CC disease, osteoarthritis, Lyme's disease, cachexia and autoimmune
 CC diseases e.g. myasthenia gravis, autoimmune diabetes and systemic
 CC lupus erythematosus. The nucleic acids are also useful for producing
 CC transgenic animals and the TANGO polypeptides themselves. Partial
 CC TANGO-128, 140, 197, 212, 213, 224, 239 sequences are useful in
 CC forensic biology, for diagnostic assays, prognostic assays,
 CC pharmacogenomics and for monitoring clinical trials. TANGO
 CC polypeptides are suitable for both prophylactic and therapeutic
 CC methods for treating a subject at risk of a disorder or having a
 CC disorder associated with aberrant TANGO expression. A wide range
 CC of cellular disorders can be treated.

XX
 SQ Sequence 2689 BP; 633 A; 739 C; 734 G; 583 T; 0 other;

alignment_scores:

Quality: 4797.50 Length: 874
 Ratio: 5.546 Gaps: 1
 Percent Similarity: 98.970 Percent Identity: 98.856

alignment_block:

US-10-044-807-2 x AAA47482 ..

Align seg 1/1 to: AAA47482 from: 1 to: 2689

1 MetGluCysCysArgArgAlaThrProGlyThrLeuLeuLeuPheLeuAl 17
 |||||
 67 ATGAATGCTGCCGTCGGGCAACTCCTGGCACACTGCTCTCTTTCTGGC 116
 17 aPheLeuLeuLeuSerSerArgThrAlaArgSerGluAlaSpArgaspG 34
 |||||
 117 TTTCTCTCTCCTCAGTTCCAGGACCGCACGCTCCGAGGAGCGCGGACG 166
 34 lYLeuTrpAspAlaTrpGlyProTrpSerGluCysSerArgThrCysGly 50
 |||||
 167 GCCTGGGATGCTGGGGCCCATGGAGTGAATGCTCACGCACCTGCGGG 216
 51 GlyClyAlaSerTyrSerLeuArgArgCysLeuSerSerLysSerCysGI 67
 |||||
 217 CGAGGGCCCTCCTACTCTCTGAGGCGGTGCTGAGCAGCAGAGCTGTGA 266
 67 uGlyArgAsnIleArgTyrArgThrCysSerAsnValAspCysProProG 84
 |||||
 267 AGGAAGAAATATCCGATACAGACATGCAGTAATGTGGACTGCCACAG 316
 84 luAlaGlyAspPheArgAlaGlnGlnCysSerAlaHisAsnAspValLys 100
 |||||
 317 AAGCAGGTGATTTCCGAGCTCAGCAATGCTCAGCTCATATGATGTCAG 366
 101 HisHisGlyGlnPheTyrGluTrpLeuProValSerAsnAspProAspAs 117
 |||||
 367 CACCATGGCCAGTTTATGAATGGCTTCCTCTGTCTAATGACCTGACAA 416
 117 nProCysSerLeuLysCysGlnAlaLysGlyThrThrLeuValValGluL 134
 |||||
 417 CCCATGTTCTCAAGTGGCCAAAGCAAGCAACCCCTGGTTGTTGAAC 466
 134 euAlaProLysValLeuAspGlyThrArgCysTyrThrGluSerLeuAsp 150
 |||||
 467 TAGCACCTAAGGCTTAGATGGTACGGTTGCTATACAGAACTCTTGAT 516
 151 MetCysIleSerGlyLeuCysGlnIleValGlyCysAspHisGlnLeuGI 167
 |||||
 517 ATGTGCATCAGTGGTTTATGCCAAATTTGTTGGCTGCGATCACCAGCTGG 566
 167 ySerThrValLysGluAspAsnCysGlyValCysAsnGlyAspGlySerT 184
 |||||
 567 AAGCACCGCTCAAGGAAGATTAACGTGGGGTCTGCAACGGAGATGGGTCCA 616
 184 hrCysArgLeuValArgGlyGlnTyrLysSerGlnLeuSerAlaThrLys 200
 |||||
 617 CCTGCCGGCTGGTCCGAGGGCAGTATAAATCCCAAGCTCTCCGCAACCAA 666

201 SerAspSpThrValValAlaIleProTyrGlySerArgHisIleArgLe 217
 |||||
 667 TCGGATGATACGTGGTTGCAATTCCTTATGGAAGTAGACATATTCGGCT 716
 217 uValLeuLysGlyProaspHisLeuTyrLeuGluThrLysThrLeuGlnG 234
 |||||
 717 TGTCTTAAAGGTCTCTGATCACTTATATCTGGAACCAAAACCTCCAGG 766
 234 lYThrLysGlyGluAsnSerLeuSerSerThrGlyThrPheLeuValAsp 250
 |||||
 767 GGACTAAGGTGAACACAGTCTCAGCTCCACAGGAACCTTCCTTGTGGAC 816
 251 AsnSerSerValAspPheGlnLysPheProAspLysGluIleLeuArgMe 267
 |||||
 817 AATTCTAGTGTGGACTTCCAGAAATTTCCAGACAAAGAGATACTGAGAAT 866
 267 tAlaGlyProLeuThrAlaAspPheIleValLysIleArgAsnSerGlyS 284
 |||||
 867 GGCTGGACCACTACAGCAGATTTTATTGTCAAGATTGCTAACTCGGGCT 916
 284 erAlaAspSerThrValGlnPheIlePheTyrGlnProIleIleHisArg 300
 |||||
 917 CCGCTGACAGTACAGTCCAGTTCACTTCTATCAACCCCTCATCCACGA 966
 301 TrpArgGluThrAspPhePheProCysSerAlaThrCysGlyGlyGlyTY 317
 |||||
 967 TGGAGGGAGACGGATTCTTTTCCITGGCTCAGCAACCTGTGGAGGAGTTA 1016
 317 rGlnLeuThrSerAlaGluCysTyrAspLeuArgSerAsnArgValValA 334
 |||||
 1017 TCAGCTGACATPCGGCTGAGTGTACGATCTGAGGAGCAACCGCTGTGGTTG 1066
 334 laAspGlnTyrCysHisTyrTyrProGluAsnIleLysProLysProLys 350
 |||||
 1067 CTGACCAATACTGTCACTATTACCCAGAGAACATCAAAACCCCAACCC 1116
 351 LeuGlnGluCysAsnLeuAspProCysProAla 361
 |||||
 1117 CTTCAGGAGTGCAACTTGGATCTTGTCCAGCCAGGAGGGTTGGCATTAT 1166
 362SerAspGlyTyrLysGlnIleMetProTyrAspLeuTyrHr 375
 |||||
 1167 TGTTCCTTAACAGTGACGGATACAGCAGATCATGCCCTTATGACCTTACC 1216
 375 isProLeuProArgTirpGluAlaThrProTirpThrAlaCysSerSerSer 391
 |||||
 1217 ATCCCTTCTCTCGTGGGAGGCCACCCCATGGACCGGTGCTCTCTCTCG 1266
 392 CysGlyGlyGlyIleGlnSerArgAlaValSerCysValGluGluAspI 408
 |||||
 1267 TGTGGGGGGGCATCCAGAGCGGGCAGTTTCTCTGTGGAGGAGGACAT 1316
 408 eGlnGlyHisValThrSerValGluGluTirpLysCysMetTyrThrProL 425
 |||||
 1317 CCAGGGGCATGTCACTTCAGTGGAGAGTGGAAATGCATGTACACCCCTA 1366
 425 ysMetProIleAlaGlnProCysAsnIlePheAspCysProLysTirpLeu 441
 |||||
 1367 AGATGCCCATCGCCGACCCCTGCAACATTTTGTACTGCCCTAATATGGCTG 1416
 442 AlaGlnGluTrpSerProCysThrValThrCysGlyGlnGlyLeuArgTy 458
 |||||
 1417 GCACAGGAGTGGTCTCGGTGCACAGTGCATGTGGCCAGGGCCCTCAGATA 1466
 458 rArgValValLeuCysIleAspHisArgGlyMetHisThrGlyCysS 475
 |||||
 1467 CCGTGTGGTCTCTTGCATCGCAATCGAGGAATGCACACAGAGGCTGTA 1516
 475 erProLysThrLysProHisIleLysGluGluCysIleValProThrPro 491
 |||||
 1517 GCCCAAAACAAAGCCACACATAAAGAGGAATGCATCGTACCCACTCCC 1566

492 CysTyrLysProLysGluLysLeuProValGluAlaLysLeuProTrpPh 508
 1567 TGTATATAAACCAGAGAACTTCCAGTCGAGGCCAAGTTCATGTT 1616
 508 eLysGlnAlaGlnGluLeuGluGlyAlaAlaValSerGluLysPro 525
 1617 CAAACAAAGCTCAAGAGCTAGAAGAGAGCTGCTGTCTCAGAGAGCCCT 1666
 525 erPheIleProGluAlaTrpSerAlaCysThrValThrCysGlyValGly 541
 1667 CGTTTCATCCAGAGGCTGTGTCGCTGCACAGTCACCTGTGTGTGGG 1716
 542 ThrGlnValArgIleValArgCysGlnValLeuLeuSerPheSerGlnSe 558
 1717 ACCCAGGTGGAATAGCTAGGTGCCAGGTGCTCTGTCTTCTCTCAGTC 1766
 558 rValAlaAspLeuProIleAspGluCysGluGlyProLysProAlaSerG 575
 1767 CGTGGCTGACCTGCCATTATGACAGAGTGTGAAGGCCCAAGCCAGCATCCC 1816
 575 lnArgAlaCysTyrAlaGlyProCysSerGlyGluIleProGluPheAsn 591
 1817 AGCGTGCTCTTATGACAGGCCCATGACGGGGGAAATTCCTGAGTTCAAC 1866
 592 ProAspGluThrAspGlyLeuPheGlyGlyLeuGlnAspPheAspGluLe 608
 1867 CCACAGCAGACAGATGGGTCTTTGTGGCTGCAGGATTCGACGAGCT 1916
 608 uTyAspTrpGluTyrGluGlyPheThrLysCysSerGluSerCysGlyG 625
 1917 GPATGACTGGAGATGAGGGGTTCACCAAGTCTCGAGTCTCTGTGGAG 1966
 625 lyGlyValGlnGluAlaValValSerCysLeuAsnLysGlnThrArgGlu 641
 1967 GAGGTGTCCAGGAGGCTGTGTGAGCTGTGAACAAACAGACTCGGGAG 2016
 642 ProAlaClnuAsnLeuCysValThrSerArgArgProProGlnLeuLe 658
 2017 CTTGCTGAGGAGAACCTGTGCTGAGCAGCGCCGCCGCCACAGCTCCT 2066
 658 uLysSerCysAsnLeuAspProCysProAlaArgTrpGluIleGlyLys 675
 2067 GAAGTCTGCAATTTGGATCCCTGCCAGCAAGGTGGAAATTTGCAAGT 2116
 675 rpSerProCysSerLeuThrCysGlyValGlyLeuGlnThrArgAspVal 691
 2117 GGAGTCCATGTAGTCTCATGTGTGGGTCTGACAGACAGAGAGCGTC 2166
 692 PheCysSerHisLeuLeuSerArgGluMetAsnGluThrValIleLeuAl 708
 2167 TTTCTGAGCCACCTGCTTTCCAGAGAGATGAATGAACAGTCATCTCTGGC 2216
 708 aAspGluLeuCysArgGlnProLysProSerThrValGlnAlaCysAsnA 725
 2217 TGATGAGCTGTGCGCAGCCCAAGCCACGACGCTCCAAGCTTGTAAAC 2266
 725 rgPheAsnCysProProAlaTrpTyrProAlaGlnTrpGlnProCysSer 741
 2267 GCTTTAATTGCCCCAGCGCTGTGACCTGTGCACAGTGGCAGCGGTGTC 2316
 742 ArgThrCysGlyGlyValGlnLysArgGluValLeuLeuCysLysGlnAr 758
 2317 AGAACGTGTGGCGGGGTGTTCAGAACGCTGAGGTCTTTTGCAGACGCG 2366
 758 gMetAlaAspGlySerPheLeuGluLeuProGluThrPheCysSerAlas 775
 2367 CATGGCTGATGGCAGCTTCTGTGAGCTTCTGTGACCTTCTGTTCAGCTT 2416
 775 erLysProAlaCysGlnGlnAlaCysLysLysAspAspCysProSerGlu 791
 2417 CAAACACCTGCTGCCAGCAGCATGCAAGAAAGATGACTGTCTCCAGCGAG 2466
 792 TrpLeuLeuSerAspTrpThrGluCysSerThrSerCysGlyGluGlyTh 808

||||| TGGCTTCTCTCAGACTGGACAGAGTGTCCACAAGCTCGGGGAGGCAC 2516
 808 rGlnThrArgSerAlaIleCysArgLysMetLeuLysThrGlyLeuSerT 825
 2517 CCAGACTCGAAGGCCCATTTGCCGAAAGATGCTGAAAAACCGGCATCTCAA 2566
 825 hrValValAsnSerThrLeuCysProProLeuProPheSerSerSerIle 841
 2567 CGTTGTCAATTCACCTGTGCCCGCCCTGCGCTTCTCTCTCTCTCCATC 2616
 842 ArgProCysMetLeuAlaThrCysAlaArgProGlyArgProSerThrLy 858
 2617 AGGCCCTGTATGCTGCGCAACCTGTGCAAGGCCGCGCGGCATCCACGAA 2666
 858 shISerProHisIleAlaIle 865
 2667 GCACAGCCCGCACATCGCGGCC 2688
 seq_name: /SIDS1/gcgdata/hold-geneseq/geneseqn-emb1/NA2000.DAT:AAA47458
 seq_documentation_block:
 ID AAA47458 standard; cdna; 2689 BP.
 XX AC AAA47458;
 XX DT 20-OCT-2000 (first entry)
 XX DE Human TANGO 224 coding sequence.
 XX KW TANGO; 128; 140; 197; 212; 213; 224; 239; modulating agent; asthma;
 KW graft versus-host diseases; rheumatoid arthritis; psoriasis;
 KW inflammatory bowel disease; septic shock; ulcerative colitis;
 KW Crohn's disease; chronic myelogenous leukemia; cancer; liver
 KW disease; Hodgkin's disease; osteoarthritis; Lyme's disease;
 KW cachexia; autoimmune disease; myasthenia gravis; autoimmune diabetes;
 KW systemic lupus erythematosus; transgenic animal; diagnosis;
 KW prognosis; prophylactic; therapeutic; human; ds.
 XX OS Homo sapiens.
 XX FH Key Location/Qualifiers
 FT CDS 1..1443
 FT /*tag= a
 FT /product= TANGO 224
 XX PN WO200039284-A1.
 XX PD 06-JUL-2000.
 XX PF 23-DEC-1999; 99WO-US31025.
 XX PR 30-DEC-1998; 98US-0223546.
 XX PA (MILL-) MILLENNIUM PHARM INC.
 XX PI Holtzman DA;
 XX DR WPI: 2000-465743/40.
 XX DR P-PSDB; AAB01425.
 XX PT Novel nucleic acid sequences encoding TANGO-128, 140, 197, 212, 213,
 XX PT 224 and 239 polypeptides useful for the treatment of asthma, rheumatoid
 XX PT arthritis, psoriasis and autoimmune diseases
 XX PS Claim 1; Fig 7; 209pp; English.
 XX CC Nucleic acids encoding TANGO polypeptides are useful as modulating
 XX CC agents for regulating cellular processes like asthma, graft
 XX CC versus-host diseases, rheumatoid arthritis, psoriasis, inflammatory
 XX CC bowel disease, septic shock, ulcerative colitis, Crohn's disease,
 XX CC chronic myelogenous leukemia, cancer, liver disease, Hodgkin's
 XX CC disease, osteoarthritis, Lyme's disease, cachexia and autoimmune

CC diseases e.g. myasthenia gravis, autoimmune diabetes and systemic
 CC lupus erythematosus. The nucleic acids are also useful for producing
 CC transgenic animals and the TANGO polypeptides themselves. Partial
 CC TANGO-128, 140, 197, 212, 213, 224, 239 sequences are useful in
 CC forensic biology, for diagnostic assays, prognostic assays,
 CC pharmacogenomics and for monitoring clinical trials. TANGO
 CC polypeptides are suitable for both prophylactic and therapeutic
 CC methods for treating a subject at risk of a disorder or having a
 CC disorder associated with aberrant TANGO expression. A wide range
 CC of cellular disorders can be treated.

XX SQ Sequence 2689 BP; 633 A; 740 C; 734 G; 582 T; 0 other;

alignment_scores:

Quality: 4775.00 Length: 874
 Ratio: 5.520 Gaps: 2
 Percent Similarity: 98.970 Percent Identity: 98.741

alignment_block:

US-10-044-807-2 x AAAA7458 ..

Align seg 1/1 to: AAAA7458 from: 1 to: 2689

```

1 MetGluCysCysArgArgAlaThrProGlyThrLeuLeuLeuPheLeuAl 17
|||||
67 ATGGAATGTCGCCGCGGGCAATCCTCGCACACATGCTCTCTTTCTGCG 116
|||||
17 aPheLeuLeuLeuSerSerArgThrAlaArgSerGluGluAspArgAspG 34
|||||
117 TTTCCGTCCTCAGTTCCAGGACCGCACGCTCCGAGGAGCAGCGGAGCG 166
|||||
34 lyLeuTrpAspAlaTrpGlyProTrpSerGluCysSerArgThrCysGly 50
|||||
167 GCCTGTGGGATGCTCGGGGCCATGGAGTGAATGCTCACGCACCTCGCGG 216
|||||
51 GlyClyAlaSerIySerLeuArgArgCysLeuSerSerIySerCysGl 67
|||||
217 GGAGGGCCCTCCCTACTCTCTGAGGGCGCTGCCTGAGCAGCAAGAGCTGTGA 266
|||||
67 uGlyArgAsnIleArgTyrArgThrCysSerAsnValAspCysProProG 84
|||||
267 AGGAAGAAATATCCGATACAGACATGCAGTAATGTGGAGTGGCCACCAG 316
|||||
84 luAlaGlyAspPheArgAlaGlnGlnCysSerAlaHisAsnAspValLys 100
|||||
317 AAGCAGGTGATTTCCGAGCTCAGCAATGCTCAGCTCATATAATGATGTCAAG 366
|||||
101 HisHisGlyGlnPheTyrGluTrpLeuProValSerAsnAspProAspAs 117
|||||
367 CACCATGGCCAGTTTATGAATGGCTTCCCTGTCTTAATGACCCCTGACAA 416
|||||
117 nProCysSerLeuLysCysGlnAlaLysGlyThrThrLeuValValGluL 134
|||||
417 CCCATGTTCACTCAAGTGCACGCAAGCAAGGAAACACCTCGTTGTGCAAC 466
|||||
134 euAlaProLysValLeuAspGlyThrArgCysTyrThrGluSerLeuAsp 150
|||||
467 TAGCACCCTAAGGCTTAGATGGTACGCGTTGCTATACAGAAATCTTTGGAT 516
|||||
151 MetCysIleSerGlyLeuCysGlnIleValIleGlyCysAspHisGlnLeuGl 167
|||||
517 ATGTGCATCAGTGGTTATGCCAAATGTGGTGCATCACCAGCTGGG 566
|||||
167 ySerThrValLysGluAspAsnCysGlyValCysAsnGlyAspGlySerT 184
|||||
567 AAGCACCCTCAAGGAACATAACTGTGGGGTCTCAACGGAGATGGGTCCA 616
|||||
184 hrCysArgLeuValArgGlyGlnTyrIyLysSerGlnLeuSerAlaThrLys 200
|||||
617 CCTGCGGGCTGGTCCGAGGCGATATAAATCCAGGCTCTCCGCAACCCAAA 666
|||||
201 SerAspAspThrValValAlaIleProTyrGlySerArgHisIleArgLe 217

```

```

|||||
667 TCGGATCATATCTGCTGGTTGCAATTCCTCTATGGAAGTAGACATATTCGCT 716
|||||
217 uValLeuLysGlyProAspHisLeuTyrLeuGluThrLysThrLeuGlnG 234
|||||
717 TGTCTTAAAGGTCCTGATCATTATATCTGGAACCAACCAACCTCCAGG 766
|||||
234 lyThrLysGlyGluAsnSerLeuSerThrGlyThrPheLeuValAsp 250
|||||
767 GGACTAAGGTGAACACAGCTCAGCTCCACAGGAACCTTCTTGTGGAC 816
|||||
251 AsnSerSerValAspPheGlnLysPheProAspLysGluIleLeuArgMe 267
|||||
817 AATTCTAGTGTGGACTTCCAGAAATTTCCAGACAAAGAGATACAGAAAT 866
|||||
267 tAlaGlyProLeuThrAlaAspPheIleValLysIleArgAsnSerGlyS 284
|||||
867 GGCCTGGACCACCTACAGCAGATTTCTTCTTGTCTCAGCAACCTGTGGAGGAGTTA 916
|||||
284 erAlaAspSerThrValGlnPheIlePheTyrGlnProIleIleHisArg 300
|||||
917 CCGCTGACAGTACAGTCCAGTTCACTTCTATCAACCCATCATCCACCGA 966
|||||
301 TrpArgGluThrAspPhePheProCysSerAlaThrCysGlyGlyGlyTy 317
|||||
967 TGGAGGAGACGGATTTCTTCTTGTCTCAGCAACCTGTGGAGGAGTTA 1016
|||||
317 rGlnLeuThrSerAlaGluCysTyrAspLeuArgSerAsnArgValValA 334
|||||
1017 TCAGCTCACATCGCTGAGTGTACGATCTGAGGAGCAACCGTGTGGTTG 1066
|||||
334 laAspGlnTyrCysHisTyrTyrProGluAsnIleLysProLysProLys 350
|||||
1067 CTGACCAATACTGTCACTATTACCAGAGAAACATCAAAACCAACCAACG 1116
|||||
351 LeuGlnGluCysAsnLeuAspProCysProAla..... 361
|||||
1117 CTTCAGGAGTGCAACTTGGATCTTGTGTCAGCCAGAGGGTTGGCATTAAT 1166
|||||
362 .....Ser..AspGlyTyrLysGlnIleMetProTyrAspLeuTyrHi 375
|||||
1167 GTTCTTAACAGTGAGCGGATACAAACAGATCATGCTTATGACCTCTACCA 1216
|||||
375 sProLeuProArgTrpGluAlaThrProTrpThrAlaCysSerSerSerc 392
|||||
1217 TCCCTTCTCCCTGGTGGAGGCCACCCCATGGACCGCGTGTCTCTCTCGT 1266
|||||
392 ysGlyGlyGlyIleGlnSerArgAlaValSerCysValGluGluAspIle 408
|||||
1267 GTGGGGGGGGCATCCAGCGGGGCGAGTTTCTGTGTGGAGGAGGACATC 1316
|||||
409 GlnGlyHisValThrSerValGluGluTrpLysCysMetTyrThrProLy 425
|||||
1317 CAGGGGCATGTCACCTCAGTGGAAAGTGGAAATGCATGTACACCCCTAA 1366
|||||
425 sMetProIleAlaGlnProCysAsnIlePheAspCysProLysTrpLeuA 442
|||||
1367 GATGCCCATCGCGAGCCCTGCAACATTTTGTACTGCCCTAAATGGCTGG 1416
|||||
442 laGlnGluTrpSerProCysThrValThrCysGlyGlnGlyLeuArgTyr 458
|||||
1417 CACAGGAGTGGTCTCCGTGCACAGTGACATGTGGCCAGGGCCTCAGATAC 1466
|||||
459 ArgValValLeuCysIleAspHisArgGlyMethHisThrGlyGlyCysSe 475
|||||
1467 CGTGTGTCTCTCATCGACCATCGAGGAATGCACACAGGAGGCTGTAG 1516
|||||
475 rProLysThrIyLysProHisIleLysGluGluCysIleValProThrProC 492
|||||
1517 CCCCCAAACAAAGCCCCACATAAAGAGGAATGCATGTACCCACCTCCCT 1566
|||||
492 ysTyrLysProLysGluLysLeuProValGluAlaLysLeuProTrpPhe 508
|||||

```

```
1567 GCTATTAACCCAAAGAGAAACTTCCAGTCAGAGGCCAAGTGGCCATGGTTC 1616
509 LysGlnAlaGlnGluLeuGluGluGlyAlaAlaValSerGluGluProSe 525
|||||
1617 AAACAAGCTCAACAGCTAGAGAAGGAGCTGCTGTGCACAGGAGCCCTC 1666
525 rPheIleProGluAlaTrpSerAlaCysThrValThrCysGlyValGlyT 542
|||||
1667 GTTCATCCCAAGAGCCCTGTCGGCCCTGCACAGTCACCTGTGGTGGGGA 1716
542 hrGlnValArgIleValArgCysGlnValLeuLeuSerPheSerGlnSer 558
|||||
1717 CCAGGTGCGAATAGTCAGGTGCAGGTGCTCTCTCTCTCTCTCAGTCC 1766
559 ValAlaAspLeuProIleAspGluCysGluGlyProLysProAlaSerG1 575
|||||
1767 GTGGCTGACCTGCTATTGACGAGTGTGAAGGCCCAAGCCAGCATCCCA 1816
575 nArgAlaCysTyrAlaGlyProCysSerGlyGluIleProGluPheAsnP 592
|||||
1817 GCGTGGCTGTATTGCAAGGCCCATGCAGCGGGGAAATTCCTGAGTTCAACC 1866
592 roAspGluThrAspGlyLeuPheGlyLeuGlnAspPheAspGluLeu 608
|||||
1867 CAGCGAGACAGATGGCTCTTTGGTGGCTGCAGGATTCGAGAGCTG 1916
609 TyrAspTrpGluTyrGluGlyPheThrLysCysSerGluSerCysGlyG1 625
|||||
1917 TATGACTGGGAGTATGAGGGGTTTCACCAAGTGTCTCGAGTCCGTGAGG 1966
625 yGlyValGlnGluAlaValValSerCysLeuAsnLysGlnThrArgGluP 642
|||||
1967 AGGTGTCAGGAGGCTGTGCTGAGTGTCTTGAACAAACACAGACTCGGAGC 2016
642 roAlaGluGluAsnLeuCysValThrSerArgArgProGlnLeuLeu 658
|||||
2017 CTGCTGAGGAGAACCTGTGCTGACACAGCGCGCGCCCCACAGCTCCTG 2066
659 LysSerCysAsnLeuAspProCysProAlaArgTrpGluIleGlyLysTr 675
2067 AAGTCTGCAATTTGGATCCTCCCAAGAGTGGGAAATTCGCAAGTG 2116
675 pSerProCysSerLeuThrCysGlyValGlyLeuGlnThrArgAspValP 692
|||||
2117 GAGTCCATGTAGTCTCACATGTGGGTGCGGCTACACAGCAGAGCGTCT 2166
692 heCysSerHisLeuLeuSerArgGluMetAsnGluThrValIleLeuAla 708
|||||
2167 TCTGCAGCACCTGCTTCCAGAGATGATGATGAACAGTCATCCCGGCT 2216
709 AspGluLeuCysArgGlnProLysProSerThrValGlnAlaCysAsnAr 725
2217 GATGAGCTGTGTCGCCAGCCCAAGCCAGCACGCTGCAAGCTGTGAACCG 2266
725 qPheAsnCysProProAlaTrpTyrProAlaGlnTrpGlnProCysSera 742
|||||
2267 CTTTAATGCCCCCAGCCCTGTACCTCGACAGTGCAGCCCGTGTCCA 2316
742 rgThrCysGlyGlyValGlnLysArgGluValLeuCysLysGlnArg 758
|||||
2317 GAACGTGTGGCGGGGTGTTCAGAAACGTGAGGTCTTTGCAAGCAGCGC 2366
759 MetAlaAspGlySerPheLeuGluLeuProGluThrPheCysSerAlase 775
2367 ATGGCTGATGGCAGCTTCCTGGAGCTTCCTGACACCTTCGTGTTTCAGTTC 2416
775 rLysProAlaCysGlnGlnAlaCysLysLysAspAspCysProSerGluT 792
|||||
2417 AAAACCTGGCTGCCAAGCAGCATGCAAGAAAGATGACTGTGCCAGGAGT 2466
792 rPheLeuSerAspTrpThrGluCysSerThrSerCysGlyGluGlyThr 808
|||||
2467 GGGTTCTCTCAGACTGGACAGAGTGTTCACAAGCTGCGGGGAAGGCACC 2516
```

```
809 GlnThrArgSerAlaIleCysArgLysMetLeuLysThrGlyLeuSerTh 825
|||||
2517 CAGACTCGAAGCGCATTTGCCGAAACATGCTGAAACCGGCATCTCAAC 2566
825 rValValAsnSerThrLeuCysProProLeuProPheSerSerSerIleA 842
|||||
2567 GGTGTCAATTCACCCCTGTGCCCGCCCTGCTTCTCTTCCTCCATCA 2616
842 rgProCysMetLeuAlaThrCysAlaArgProGlyArgProSerThrLys 858
|||||
2617 GGCCTGTATGCTGGCAACCTGTGCAAGGCCCGGGCGGCATCCACGAAG 2666
859 HisSerProHisIleAlaAla 865
|||||
2667 CACAGCCCGCACATCGCGGCC 2687

seq_name: /SIDS1/gcgdata/hold-geneseq/geneseq-n-emb1/NA2001A.DAT:AAD14364
seq_documentation_block:
ID AAD14364 standard; cdna; 5076 BP.
XX
AC AAD14364;
XX
DT 01-NOV-2001 (first entry)
XX
DE Novel human protein (NHP) cdna #1.
XX
KW Novel human protein; NHP; thrombospondin; gene therapy; cancer;
cytostatic; antisense therapy; angiogenesis; biological disorder; ss.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT CDS 1..5076
FT FT /*tag= a
FT FT /product= *Novel human protein*
XX
PN WO200161011-A2.
XX
PD 23-AUG-2001.
XX
PF 15-FEB-2001; 2001WO-US05290.
XX
PR 17-FEB-2000; 2000US-0183282.
XX
PA (LEXI-) LEXICON GENETICS INC.
XX
PI Donoho G, Scoville J, Turner CA, Friedrich G, Zambrowicz B;
Sands AT;
XX
DR WPI; 2001-514776/56.
DR P-PSDB; AAE07863.
XX
PT An isolated nucleic acid encoding a thrombospondin useful as a
hybridization probe and gene therapy treatments of cancer -
XX
PS Claim 1; Page 29-30; 56pp; English.
XX
CC The invention relates to novel human polynucleotides encoding proteins
that share sequence similarity with animal proteins having thrombospondin
repeats. NHPs are expressed in human cell lines, pituitary, lymph nodes,
prostate, testis, adrenal gland, uterus, foetal kidney, foetal lung and
gene trapped human cells. NHPs are thrombospondins useful for treating
biological disorders involving angiogenesis, cancer and development and
also in pharmacogenomic applications. NHPs are useful as a hybridisation
probe for screening libraries, assessing gene expression patterns and
also in gene therapy. Proteins having thrombospondin repeats act as
receptors, secreted extracellular matrix proteins and proteases. The
present sequence is a cdna encoding novel human protein (NHP).
XX
SQ Sequence 5076 BP; 1384 A; 1227 C; 1338 G; 1127 T; 0 other;
```

```
alignment_scores:
  Quality: 3956.50      Length: 1775
  Ratio: 3.248          Gaps: 31
  Percent Similarity: 68.620  Percent Identity: 43.324

alignment_block:
US-10-044-807-2 x AAD14364 ..
Align seg 1/1 to: AAD14364 from: 1 to: 5076

22 SerSerArgThrAlaArgSerGluGluAspArgAspGlyLeuTrpAspAl 38
:::|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
190 ACCTCAAGAAACACTCGTTCAGATGAAGCAAAAGATCGCAACTGGGATGC 239
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
38 atpGlyProTrpSerGluCysSerArgThrCysGlyGlyGlyAlaSerT 55
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
240 TTGGGGCGACTGGAGTGACTGCTCCCGGACCTGTGGGGAGGAGCATCAT 289
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
55 yrSerLeuArgCysLeuSerSerLysSerCysGluGlyArgAsnIle 71
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
290 ATTCTCTGGGAGATGTTGACTTGAAGGAATGCTGAAGGGCAGAATTT 339
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
72 ArgTyrArgThrCysSerAsnValAspCysProGluAlaGlyAspPh 88
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
340 CGGTACAGACATGCGACCAATCATGACTGCCCTCCAGATCGCAAGATTT 389
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
88 eArqAlaGlnClnCysSerAlaHisAsnAspValLysHisGlyGlnP 105
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
390 CAGAGCCACGACGTCTCAGCCTACAAATGATGCCAGTATCAGGGGCATT 439
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
105 heTyrGluTrpLeuProValSerAsnAspProAspAsnProCysSerLeu 121
:::|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
440 ACTATGAATGGCTCCACGATATAATGATCTGCTGCCCGGTGCACTC 489
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
122 LysCysGluAlaValGlyThrThrLeuValValGluLeuAlaProLysVa 138
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
490 AAGTGATCATGCAAGGACAAAACCTTGGTGGTGGAGTGGCACTTAAGGT 539
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
138 lLeuAspGlyThrArgCysTyrThrGluSerLeuAspMetCysIleSerG 155
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
540 ACTGGATGGAACTCGTGCAACACGGACCTCTTGGACATGTATCAGTG 589
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
155 lYLeuCysGlnIleValGlyCysAspHisGlnLeuGlySerThrValLys 171
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
590 GCATCTGTACAGGACGTGGGTGGATCGGCACTGGGAACAATGCCAAG 639
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
172 GluAspAsnCysGlyValCysAsnGlyAspGlySerThrCysArgLeuVa 188
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
640 GAGGACAACTGTGGATGCTGTGCGCGGATGGCTCCACCTGCGAGCTTGT 689
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
188 lArGlyGlnTyrLysSerGlnLeuSerAlaThrLysSerAspThrY 205
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
690 ACGGGACAATCAAGTCAACAGTTCCTCTCTGAAAGAAAGAGAGAAAATG 739
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
205 alValAlaIleProTyrGlySerArgHisIleArgLeuValLeuLysGly 221
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
740 TAAATGCTGTCTTGGGAGTCGAAGTGTGAGAAATTACAGTCAAGGA 789
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
222 proAspHisLeuTyrLeuGluThrLysThrLeuGlnGlyThrLysGlyG 238
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
790 CTTGCCCACTCTTATTTGATGATCAAAACACTTCAGGAAGCAAGGAGA 839
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
238 uAsnSerLeuSerSerThrGlyThrPheLeuValAspAsnSerSerVala 255
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
840 ACACAGCTTTAAACAGCCCGCGCTCTTGTGCTAGAAAAACACACAGTGG 889
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
255 spPheGlnLysPheProAspLysGluIleLeuArgMetAlaGlyProLeu 271
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
890 AATTTACAGGGGCTCCGAGAGCAACTTTTAAAGATTCAGGACCTCTG 939
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
272 ThrAlaAspPheIleValLysIleArgAsnSerGlySerAlaAspSerTh 288
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
```

```
940 ATGGCTGATTTTCATCTTCAAGACCAGGTACACTGCAGCCAAAGACAGCGT 989
288 rValGlnPheIlePheTyrGlnProIleIleHisArgTyrArgGluThrA 305
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
990 GGTTCAGTTCTTCTTTACAGCCCATCAGTCATCAGTGGAGACAAACTG 1039
305 spPhePheProCysSerAlaThrCysGlyGlyGlyTyrGlnLeuThrSer 321
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1040 ACTTCTTCCCTGCACGTGACGTGTGGAGGAGGTATCAGTCAATCT 1089
322 AlaGluCysTyrAspLeuArgSerAsnArgValValAlaAspGlnTyrCy 338
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1090 GCTGAATGTGGATATCCGCTTCAAGAGGTAGTTCTTCCGACCAATTATG 1139
338 sHisTyrIlyrProGluAsnIleLysProLysProLysLeuGlnGluCysA 355
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1140 TCACTACTACCCIGAAATGTAAACCAACCAACCAAACTGAAGGAATGCA 1189
355 snLeuAspProCysProAlaSerAspGlyTyrLysGlnIleMetProTyr 371
:::|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1190 GCATGGATCCCTGCCCATCAAGTGTGGATTTAAAGAGATAATGCCCTAT 1239
372 AspLeuTyrHisProLeuProArgTyrGluAlaThrProTrpThrAlaCy 388
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1240 GACCACTTCCAACCTCTCCCTCGCTGGAAATTAATCTTGGACTGCATG 1289
388 sSerSerSerCysGlyGlyGlyIleGlnSerArgAlaValSerCysValG 405
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1290 TTCGGTCTCTGTGGAGGAGGATTCAGAGACGGAGCTTTTGTGTGTAG 1339
405 luGluAspIleGlnGlyHisValThrSerValGluGluTrpLysCysMet 421
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1340 AGGAATCCATGCATGGAGAGATATTCAGGTGGGAAGATGGAAGTGCATG 1389
422 TyrThrProLysMetProIleAlaGlnProCysAsnIlePheAspCysPr 438
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1390 TACCACCCAAACCCCAAGGTATGCAACTGTAACTGTGTGTGTGTGTC 1439
438 oLysTrpLeuAlaGlnGluTrpSerProCysThrValThrCysGlyGlnG 455
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1440 CAAGTGGATIGCCATGGAGTGTCTCAGTCACAGTGACITGTGGCCGAG 1489
455 lYLeuArgTyrArgValValLeuCysIleAspHisArgGlyMetHisThr 471
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1490 GGTATACGGTACCGGGTGTCTGTGTATTAACCAACCGCGAGAGCATGT 1539
472 GlyGlyCysSerProLysThrLysProHisIleLysGluGluCysIleVa 488
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1540 GGGGCTGCATCCCAACTGAAGTTACACATCAAGAAAGAAATGTGTCA 1589
488 lProThrProCysTyrLysProLysGluLysLeuProValGluAlaLysL 505
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1590 TCCCATCCGCTGTATAAACCAAAAGAAAAAGTCCAGTGGAGCAAAAT 1639
505 euProTrpPheLysGlnAlaGlnGluLeuGluGlyAlaAlaValSer 521
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1640 TGCCTTGGCTGAAACAGCACAAAGAACTAGAAAGACCAAGAAATAGCA 1689
522 GluGluProSerPheIleProGluAlaTrpSerAlaCysThrValThrCy 538
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1690 GAAGAACCAAGTTCATTCAGAACCCCTGGTCAAGCTGCAGTACCACGTG 1739
538 sGlyValGlyThrGlnValArgIleValArgCysGlnValLeuLeuSerP 555
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1740 TGGGCCAGGTGTGAGGTCCCGAGGTGAAGTGCCTGTGTCTCTCATAT 1789
555 heSerGlnSerValAlaAspLeuProIleAspGluCysGluClyProLys 571
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1790 TCAGCAGACTGAGACTGAGCTGCCCGAGGAAGAGTGTGAAGGCCCAAG 1839
572 ProAlaSerGlnArg.....AlaCysTyrAlaGlyProCy 583
:::|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1840 CTGCCCAACCGACCGCTCTCTCTGGAACGATGTGATGAGAGCCCGGC 1889
```

583 sSerGlyLeuIleProGluPheAsnProAspGluThrAspGlyLeuPheG 600
1890 CTCGAGAGCTACATCCCTCCCT..... 1917
600 lyGlyLeuGlnAspPheAspGluLeuThrAspTrpGluThrGluGlyPhe 616
1918GAGGACAGTACAGACACCTACACCTGGAGTACGCTGGGTTCC 1959
617 ThrLysCysSerGluSerCysGlyGlyValGlnGluAlaValValse 633
1960 ACCCTTGCACAGCAACATCTTGGGAGGCCATCAAGAACCCATAGCAGT 2009
633 rCysLeuAsnLysGlnThrArgGluProAlaGluAlaAsnLeuCysValt 650
2010 GTCTTACATATCAGACCCAGCAGACAGTCAATGACAGCTGTGTGATA 2059
650 hrSerArgArgProGlnLeuLeuLysSerCysAsnLeuAspProCys 666
2060 TGTCTCACCGCTCTCCAGCCATGAGCGGCCCTGTACACAGAGCCCTGT 2109
667 ProAlaArgTrpGluLeuLysTrpSerProCysSerLeuThrCysG1 683
2110 CCCCCAGGTGGCATGTGGGCTCTTGGGGCCCTGCTCAGCTACCTGTGG 2159
683 yValGlyLeuGlnThrArgAspValPheCysSerHisLeuLeuSerArgG 700
2160 AGTGGAAATTCAGACCCGAGATGTACTGCTGCTGCAC..... 2196
700 luMetAsnGluThrValIleLeuAlaAspGluLeuCysArgGlnProLys 716
2197 ..CCAGGGAGACCCCTGCCCTCTCGAGGAG...TGCCGAGATGAAG 2241
717 ProSerThrValGlnAlaCysAsnArgPheAsnCysProProAlaTrpTy 733
2242 CCCCAGCTTTACAGCATGCAATGATTGACTGCTCCCTGGCTGGCA 2291
733 rProAlaGlnTrpGlnProCysSerArgThrCysGlyGlyValGlnL 750
2292 CATGGAAGATGGCAGCAGTGTTCAGGACTGTGTGGGGGGAACTCAGA 2341
750 ysArgGluValLeuCysLysArgMetAlaAspGlySerPheLeuGlu 766
2342 ACAGAGAGTCACTCTCGCAGCTGCTACGAGTGGCAGCTTTTGAAT 2391
767 LeuProGluThrPheCysSerAlaSerLysProAlaCysGlnAlaCy 783
2392 CTCTCAGATGAATGTGCCAAGGACCCAGGCATGCTCTCACAAAGTCTG 2441
783 sLysLysAspCysProSerClnTrpLeuLeuSerAspTrpThrGluC 800
2442 TCCGAGCAGACTGTCTCCACATTTACTGTGGGAGACTGGTCCAAGT 2491
800 ysSerThrSerCysGlyGluGlyThrGlnThrArgSerAlaIleCysArg 816
2492 GTTCTGTCAAGTGTGTGTGATTCAGAGAGAAAGCAGGTGTGTCAA 2541
817 LysMetLeuLysThrGlyLeuSerThrValValAsnSerThrLeuCysPr 833
2542 AGGCTGGCAGCCAAAGGTCGGGCATCCCTCCAGTGAGATGATGTGCAG 2591
833 oProLeuProPheSerSerIleArgProCysMetLeuAlaThrCysA 850
2592 GATCTACCAAGGTTCCCTCTGTGAAGATCTTGCCAGATGCCTAGTGCA 2641
850 laArg...ProGlyArgProSerThrLys.....HisSerProHis 862
2642 GTAATCAATCAAGATGAGACACAAACTTGGTCAGCAGGTTCGCAG 2691
863 IleAlaAlaAlaArgLysValTyrIleGlnThrArgGlnArgLysLe 879
2692 ATCCTCAGTGTCCAGAGACTTACATTCAGACAAAGGAAGAGAGCGTAT 2741

879 uHisPheValValGlyGlyPheAlaTyrLeuLeuProLysThrAlaValV 896
2742 TAACCTGACCATTTGGTAGCAGAGCTATTTCCTGCCAACACATCCGTGA 2791
896 alLeuArgCysProAlaArgArgValArgLysProLeuIleThrTrpGlu 912
2792 TTATTAAGTGCCTGCGCGAGGATTCCAGAAATCTCTGATCCAGCTGGGAG 2841
913 LysAspGlyGlnHisLeuSerSerThrHisValThrValAlaProPh 929
2842 AAGATGGCGTTCCTGCAAGACTCAACCGCTTGGCATCACCAGATC 2891
929 eGlyTrpLeuLysIleHisArgLeuLysProSerAspAlaGlyValTyr 946
2892 AGGCTCACTAAATCCACGCTCTTGTGCCCCCGACATCGCGCTGTACC 2941
946 hrCysSerAlaGlyProAlaArgGluHisPheValIleLysLeuIleGly 962
2942 GGTGCTATGCGGCTCTGCACAGAAACAGTTGTCTCAAGCTCATTTGT 2991
963 GlyAsnArgLysLeuValAlaArgProLeuSerProArgSerGluGluG 979
2992 ACTGACACCGGCTCATCGCAGCCCA..... 3018
979 uValLeuAlaGlyArgLysGlyProLysGluAlaLeuGlnThrHisL 996
3019GCCCTCAGGAGCCTATGAGGGAATATCTGGGATGGACACACA 3061
996 yHisGlnAsnGlyIlePheSerAsnGlySerLysAlaGluLysArgGly 1012
3062 GCGAAGCAATAGTTTGGGAGTCACTGCGCAAA.....ATGAGGCAA 3105
1013 LeuAlaAlaAsnProGlySerArgTyr.....AspAspLeuValSerAr 1027
3106 ATGTGGNATAACAAAAGACCTTATCTCGATGATGACCATATTAGTAA 3155
1027 g.....LeuLeuGluGlnGlyTrpProGlyGluL 1038
3156 CCAGCCTTTCTTGAGAGCTCTGTAGGCCACTGCAGCAATCTCGAGGAA 3205
1038 euLeuAlaSerTrpGluAlaGlnAspSerAlaGluArgAsnThrThrSer 1054
3206 GCACCAACTCTCTGGGAGTTGAAGAAATAGCAGTTTGAAGCAGCAGTTAA 3255
1055 GluGluAspProGlyAlaGluGlnValLeuLeuHisLeuProPheThrMe 1071
3256 CAA.....GGACCA.....TATAGCAT 3272
1071 tValThrGluGlnArgArgLeuAspAspIleLeuGlyAsnLeuSerGlnG 1088
3273 GGATACAGCCAG.....TTTGATGAGCTGATAAGAAACATGAGTCAGC 3316
1088 lnProGlu.....GluLeuArgAspLeuTyrSerLysHisLeuValAla 1102
3317 TCATGAAACCGGAGAGGTGACGCGATGATCTTGCCTCCAGCTGATATAT 3366
1103 GlnLeuAlaGlnGluIlePheArgSerHisLeuGluHisGlnAspThrLe 1119
3367 CAGCTGGTGGCCCAATTAGCCCAAGGACACCCCAACACATGACAG..... 3411
1119 uLeuLysProSerGluArgArgThrSerProValThrLeuSerProHisL 1136
3412 .TGCGGGGCATCCAGAAAGAGACACCTCT..... 3441
1136 yHisValSerGlyPheSerSerSerLeuArgThrSerSerThrGlyAsp 1152
3442GCTGCTCAGCTCAGA.....GGGAA 3462
1153 AlaGlyGlyLysArgArgProHisArgLys..... 1163
3463 ACAGGAGTGTGCCCAAGCTCGCATGCAAAAACTCAGGCAAGCTGAC 3512
1164ProThrIleLeuArgLysIleSerAlaAlaGlnG 1175

```

3513 ATTCAAGCCCAAGGACCTGTTCTCATGAGCAA...AGCCAACCTCCCT 3559
1175 lnLeuSerAlaSerGluValVal...ThrHisLeuGlyGlnThrValAla 1190
3560 CAATTTCAATTAATAATAAATAAATCCAGGATTCAGAAATACAGATATAC 3609
1191 LeuAlaSerGlyThrLeuSerValLeuLeuLeuHisCysGluAlaIleGlyHI 1207
3610 ATTACAAAAGGACAGAGGTCATATATACCTGTGTGACCTTATTACCCC 3659
1207 sProArgProThrIleSerTrpAlaArgasnGlyGluGluValGlnPheS 1224
3660 CAGTGAGGCCACATATACATGGACCAAGGATGGAACCTTGTTCACAGCCCT 3709
1224 erAspArgileLeuLeuGlnProAspAspSerLeuGlnIleLeuAlaPro 1240
3710 CAGTAAATAATTTGGATGGAACTGGGAAGATACAGATACAGAAATCCT 3759
1241 ValGluAlaAspValGlyPheThrCysAsnAlaThrAsnAlaLeuGI 1257
3760 ACAAGAAAGAACAGGATATATGAATGTTCTGTAGCTAATCATICTTGG 3809
1257 yTrpAspSerValSerIleAlaValThrLeuAlaGlyLysProLeuValL 1274
3810 TTCAGATGTGGAAGTCTTCTGTGCTGTATGCAGAGGCACCTGTCTATCT 3859
1274 yThrSerArgMetThrValIleAsnThrGluLysProAlaValThrVal 1290
3860 TGCTGTGGAAGAAATATACCAACACAGGACCAACCATCTGTCTGCTT 3909
1291 AspIleGlySerThrIleLysThrValGlnGlyValAsnValThrIleAs 1307
3910 GTGGTTGGAGCATCTGGAGGAGCCCTTGGACCAACGTGCAATCCG 3959
1307 nCysGlnValAlaGlyValProGluAlaGluValThrTrpPheArgAsnL 1324
3960 ATGTCCTGTAAAAGGTGCTCCCTCAGCCTAATATAACTTGGTTGAAGAG 4009
1324 ysSerLysLeu...GlySerProHisHisLeuHisGlyLysSerLeuLeu 1339
4010 GAGGATCTCTGAGTGGCAAGTTTCCTTGCTTTTCAATGATCCCTCTTG 4059
1340 LeuThrAsnValSerSerAspGlnGlyLeuTrpSerCysArgAlaAl 1356
4060 TTGCAGAAATGTTCCCTTGAATATGAAGAACCTTACGCTGCTGATAGCC 4109
1356 aAsnLeuHisGlyGluLeuThrGluSerThrGlnLeuLeuIleLeuAspP 1373
4110 CAATGCTTGGAAAGCAGTGGCAACATCTGACTCCACTTGTCTGGAAC 4159
1373 roProGlnValProThrGlnLeuGluAspIleArgAla..... 1385
4160 GAAGATGCCAGAGAGTAGAATCGTATTTCGAAGGACATACAAAGTAC 4209
1386 LeuLeuAlaAlaThrGlyProAsnLeuProSerValLeuThrSerProLe 1402
4210 ATTCTCAGGCAACCAACACAGTACCAACCAACAGC..... 4242
1402 uGlyThrGlnLeuValIleLeuAspProGlyAsnSerAlaLeuLeuGlyCysP 1419
4243 .....AATGACCCA..... 4251
1419 roIleLysGlyHisProValProAsnIleThrTrpPheHisGlyGlyGln 1435
4252 .....ACAGGAGAACCCCGCCT..... 4269
1436 ProIleValThrAlaThrGlyLeuThrHisHisIleLeuAlaAlaGlyGI 1452
4269 ..... 4269
1452 nIleLeuGlnValAlaAsnLeuSerGlyGlySerGlnGlyGluPheSerC 1469

```

```

4269 ..... 4269
1469 ysLeuAlaGlnInAsnGluAlaGlyValLeuMetGlnLysAlaSerLeuVal 1485
4269 ..... 4269
1486 IleGlnAspTrpTrpTrpSerValAspArgLeuAlaThrCysSerAlaSe 1502
4270 ...CAAGAGCCTTTTGGAGCCTGTAACTGGTCACATTTGTTCTGCCAC 4316
1502 rCysGlyAsnArgGlyValGlnGlnProArgLeuArgCysLeuLeu...A 1518
4317 CTGTGGTCATTTGGGACCCCGCATTCAGAGACCCAGTCGTGTATGCCCA 4366
1518 snSerThrGluValAsnProAlaHisCysAlaGlyLysValArgProAla 1534
4367 ATGGCAGGAAGTGAGTGAGGCCCTGTGTGATCACCTCCAGAACCCACTG 4416
1535 ValGlnProIleAlaCysAsnArgArgAspCysProSerArgTrpMetVa 1551
4417 GCTGGGTTTGGCCCTGTAACTCCGGGACTGCCGAGGAGTGGTTCAC 4466
1551 lThrSerTrpSerAlaCysThrArgSerCysGlyGlyValGlnThrA 1568
4467 AAGTGTGTGTTCACAGTCTGTCTGTTCGGTGAAGGATACACAGATC 4516
1568 rgArgValThrCysGlnLysLeuLysAlaSerGlyIleSerThrProVal 1584
4517 GGCAGGTGACGTGCAAGCGGCAAAAGCAATGGAACCTGTGCAGGTGGTG 4566
1585 SerAsnAspMetCysThrGlnValAlaLysArgProValAspThrGlnAl 1601
4567 TCTCAAGAGCATGTGCCCT...AAAGACCGGCTCTGGGAAGAAAC 4613
1601 acYsAsnGlnGlnLeuCysValGluTrpAlaPheSerSerTrpGlyGlnC 1618
4614 ATGTTTTGGTCATCCATGTGTCACTGG.....GAACCGAGGAACCGT 4657
1618 ysAsnGlyProCysIleGlyProHisLeuAlaValGlnHisArgGlnVal 1634
4658 GTCTGGACGTTGATGGCGGTGCTGTGAGGATGCAGCAGCGTCACACA 4707
1635 PheCysGlnThrArgAspGlyIleThrLeuProSerGluGlnCysSerAl 1651
4708 GCTTGTCAACACACAGCTCT.....GACTCCAACCTGTGATGA 4745
1651 aLeuProArgProValSerThrGlnAsnCysTrpSerGluAlaCysSerV 1668
4746 CAGAAAGAGACCCACCITTAAGAGGAACTGCACATCAGGGGCGCTGTGATG 4795
1668 alHisTrpArgValSerLeuTrpThrLeuCysThrAlaThrCysGlyAsn 1684
4796 TGTGTGGCACACAGGCCCTTGGAAAGCCTGTACAGCAGCGCTGTGGCAG 4845
1685 TyrGlyPheGlnSerArgValGluCysValHisAlaArgThrAsnLy 1701
4846 ...GGTTTCCAGTCTCGAAAGTCGACTGTATCCACACAGGAGTGGAA 4892
1701 sAlaValProGluHisLeuCysSerTrpGlyProArgProAlaAsnTrpG 1718
4893 ACCTGTGCCCAAGAGACACTGTGTACAGAAAAGAACCAATTTCTCTGCC 4942
1718 lnArgCysAsnIleThrProCysGluAsnMetGluCysArgAspThrThr 1734
4943 GGCACGTGT...CTTGGGCCCTCCTGTGTATAGAGACTGCACAGACACA 4989
1735 ArgTrpCysGluLysValLysGlnLeuLysLeuCysGlnLeuSerGlnPh 1751
4990 CACTACTGATGTTGTAAACATCTTAATTGTGTTCTCTAGACCCGCTA 5039
1751 eLysSerArgCysCysGlyThrCys 1759
5040 CAAACAAAGGTGCTGCCAGTCATGT 5064

```

seq_name: /STDS1/gcgdata/hold-geneseq/geneseq-emb1/NA2001A.DAT:AA17757

seq_documentation_block:

ID AAD17757 standard; DNA; 7260 BP.

XX AAD17757;

XX 10-DEC-2001 (first entry)

XX Human novel KIAA1233-like protein, NOV-2a encoding DNA.

XX Human: NOV-X protein; KIAA1233-like protein; STE20-like protein; tumour;
 KW trypsin inhibitor-like protein; gene therapy; haematopoietic; illness;
 KW immunological disorder; neurodegenerative disorder; Alzheimer's disease;
 KW Parkinson's disease; immunomodulatory; pharmacogenomic; haemostatic;
 KW human immunodeficiency virus; HIV; fertility disorder; neuroprotective;
 KW cytostatic; nootropic; anti-infertility; cancer; chromosome 15;
 KW NOV-2a protein; ds.

XX Homo sapiens.

XX Key Location/Qualifiers

FT 5'UTR 1..135

FT /*tag= a

FT 136..5211

FT /*tag= b

FT /*product= "Human novel KIAA1233-like protein, NOV-2a"

FT sig_peptide 136..213

FT /*tag= c

FT mat_peptide 214..5208

FT /*tag= d

FT /*product= "Human mature novel KIAA1233-like protein,

FT NOV-2a"

FT 5212..7260

FT /*tag= e

FT WO200162928-A2.

XX 30-AUG-2001.

XX 26-FEB-2001; 2001WO-US06151.

XX 25-FEB-2000; 2000US-0184951.

PR 28-FEB-2000; 2000US-0185548.

PR 01-MAR-2000; 2000US-0185967.

PR 18-APR-2000; 2000US-0197723.

PR 27-APR-2000; 2000US-0199957.

PR 23-FEB-2001; 2001US-0789390.

XX (CURA-) CURAGEN CORP.

XX Vernet CAM, Fernandes E, Shinkets RA, Macdougall J, Spaderna SK;

PI WPI; 2001-582051/65.

XX P-PSDB; AAE10609.

XX New isolated KIAA1233-like, STE20-like, or trypsin inhibitor-like
 PT polypeptide for diagnosing and treating pathological disorders, such as
 PT Parkinson's disease and for use in pharmacogenomics -
 XX Claim 9; Page 20-22; 189pp; English.

XX The invention relates to novel human polypeptides referred as NOV-X
 CC and their corresponding nucleic acid sequences. NOV-X collectively
 CC include NOV-1, NOV-2a and NOV-2b which are novel KIAA1233-like
 CC polypeptides, NOV-3a, NOV-3b, NOV-3c and NOV-3d which are novel
 CC STE20-like polypeptides and NOV-4a, NOV-4b, NOV-4c, NOV-4d and NOV-4e
 CC which are novel trypsin inhibitor-like polypeptides. NOV-X is used to
 CC identify a potential therapeutic agent that can modulate its activity
 CC and can be used for treating a pathology related to aberrant expression
 CC or aberrant physiological interactions of NOV-X. NOV-X or its DNA is
 CC used to determine the presence or predisposition to a disease associated

CC with altered levels of NOV-X. NOV-X, its DNA and its antibody are used
 CC to treat or prevent a pathology associated with NOV-X. The pathological
 CC states that can be treated or prevented are haematopoietic, cancer,
 CC immunological, tumour, neurodegenerative (e.g. Alzheimer's and
 CC Parkinson's disease), human immunodeficiency virus (HIV) illness and
 CC fertility disorders. NOV-X and its DNA are used in pharmacogenomics for
 CC predictive medicine. NOV-X DNA is used in gene therapy. The present
 CC sequence is a DNA encoding human novel KIAA1233-like protein, NOV-2a.
 CC NOV-2a gene is located at chromosome 15.

XX Sequence 7260 BP: 2022 A; 1593 C; 1811 G; 1834 T; 0 other;

alignment_scores:

Quality: 3953.50 Length: 1775

Ratio: 3.246 Gaps: 31

Percent Similarity: 68.620 Percent Identity: 43.268

alignment_block:

US-10-044-807-2 x AAD17757 ..

Align seg 1/1 to: AAD17757 from: 1 to: 7260

22 SerSerArgThrAlaArgSerGluGluAspArgAspGlyLeuTrpAspAl 38

325 ACCTCAGCAAAACACTCGTTTCAGATCAACACAAAGATGCCAAGTGGGATGC 374

38 aTrGlyProThrPpSerGluCysSerArgThrCysGlyGlyGlyAlaSerT 55

375 TTGGGGCGACTGAGAGTGACTGCTCCCGGACCTGTGGGGGAGGAGCATCAT 424

55 YTSerLeuArgArgCysLeuSerSerLysSerCysGluGlyArgAsnIle 71

425 ATTCCTCGCGAGATGTTTGACTGGAAGGAATTGTGAAGGCGCAGAACATT 474

72 ArgTyrArgThrCysSerAsnValAspCysProGluAlaGlyAspPh 88

475 CGGTACAAGACATGCAGCAATCATGACTGCTCCAGATGCAGATATTCAGGGCAT 524

88 eArgAlaGlnGlnCysSerAlaHisAsnAspValLysHisGlyGlnP 105

525 CAGAGCCCGCAGTGCCTCAGCTCAATGATGCTCAGTATCAGGGCAT 574

105 heTyrGluTrpLeuProValSerAsnAspProAspAspProCysSerLeu 121

575 ACTATGAATGGCTTCCAGATATATGATCTGCTGCTGCCCTGTGCACTC 624

122 LysCysGlnAlaLysGlyThrThrLeuValValGluLeuAlaProLysVa 138

625 AAGTCTCATGCAACAGGACAAAACTTGGTGTGGAGCTGGCACCTAAGGT 674

138 lLeuAspGlyThrArgCysTyrThrGluSerLeuAspMetCysIleSerG 155

675 ACTGGATGGAACTCGTTGCCAACACGCTCCTTGGACATGCTATCAGTG 724

155 lyLeuCysGlnIleValGlyCysAspHisGlnLeuGlySerThrValLys 171

725 GCATCTGTCAGGAGTGGCTGCGATCGCAACTGGGAGCAATGCCAAG 774

172 GluAspAsnCysGlyValCysAsnGlyAspGlySerThrCysArgLeuVa 188

775 GAGGACAACTGTGAGTCTGTGCCGGGATGGCTCCACCTCGAGGCTTGT 824

188 lArgGlyGlnTyrLysSerGlnLeuSerAlaThrLysSerAspAspThrV 205

825 ACGGGGCAATCAAAAGTCACACGCTTCTCTCGAAAAAAGAGAGAAAATG 874

205 alValAlaIleProTyrGlySerArgHisIleArgLeuValLeuLysGly 221

875 TAATTGCTGTCTCTTGGGAGTGGAGTGTGAGAAATACAGTGAAGGA 924

222 ProAspHisLeuTyrLeuGluThrLysThrLeuGlnGlyThrLysGlyGl 238

[illegible]

538	sglyValGlyThrGlnValArgIleValArgCysGlnValLeuLeuSerP	555
1875	TGGCGCGGGTGTGCAGGTCCGTGAGGTGAAGTGCCTGCTCTCCTC	1924
555	heSerGlnSerValAlaAspLeuProIleAspGluCysGluGlyProLys	571
1925	TCCAGCAGACTGAGACTGAGCTGCCGAGGAAGAGTGTGAAGGCCCCAG	1974
572	ProAlaSerGlnArg.....AlaCysTyraIaGlyProCys	583
1975	CTGCCACCGAAGCGCCCTGCTCTGGAAGCATGTGATGAGAGCCCGGC	2024
583	sSerGlyGluIleProGluPheAsnProAspGluThrAspGlyLeuPheC	600
2025	CTCCCGAGAGCTAGACATCCCTCTCCCT.....	2052
600	LyGlyLeuGlnAspPheAspGluLeuTyraSpTrpGluTyrglyPhe	616
2053GAGCAGACTGAGCAGCTTACGACTGGAGTACGCTGGGTTC	2094
617	ThrLysCysSerGluSerCysGlyGlyValGlnGluAlaValValse	633
2095	ACCCCTTGACAGCAACATGCGTGGAGGCCATCAAGAAGCCATAGCAGT	2144
633	rCysLeuAsnLysGlnThrArgGluProAlaGluAsnLeuCysValt	650
2145	GTGCTTACATATCCAGACCCAGACAGTCAATGACAGCTGTGTGATA	2194
650	hrSerArgArgProGlnLeuLeuLysSerCysAsnLeuAspProCys	666
2195	TGCTCCACGGTCTCCAGCCATGAGCCAGGCTGTAAACAGAGCCCTGT	2244
667	ProAlaArgTrpGluIleGlyLysTrpSerProCysSerLeuThrCysGl	683
2245	CCCCCAGGTGGCATGTGGCTCTTGGGGGCCCTGCTCAGCTACCTGTG	2294
683	yValGlyLeuGlnThrArgAspValPheCysSerHisLeuLeuSerArg	700
2295	AGTTGGAAATCAGACCCAGATGTGTACTGCCCTGCAC.....	2331
700	luMetAsnGluThrValIleLeuAlaAspGluLeuCysArgGlnProLys	716
2332	..CCAGGGGAGCCCTGCCCTCCTGAGGAG...TGCCGAGATGAAGAA	2376
717	ProSerThrValGlnAlaCysAsnArgPheAsnCysProAlaTrpTy	733
2377	CCCCATGCTTTACAGATGCAATCAGTTTACTGCTCCCTCTGCTGGCA	2426
733	rProAlaGlnTrpGlnProCysSerArgThrCysGlyGlyValGlnL	750
2427	CATTGAAGAATGGCAGCAGTGTTCAGCAGCTTGTGGCGGGGCACTCAG	2476
750	ysArgGluValLeuCysLysGlnArgMetAlaAspGlySerPheLeuGlu	766
2477	ACAGAAGATCACTCTGCGGAGCTGTAAACGGATGGCAGCTTTTGAAT	2526
767	LeuProGluThrPheCysSerAlaSerLysProAlaCysGlnGlnAlaCy	783
2527	CTCTCAGATGAATTGTGCCAAGGCCCAAGGCATGCTCTCAGAGTCCGT	2576
783	sLysLysAspCysProSerGluTrpLeuLeuSerAspTrpThrGluC	800
2577	TGCCAGGACAGACTGTCTCCACATTTACGTGTGGAGACTGGTCGAAT	2626
800	ysSerThrSerCysGlyGluGlyThrGlnThrArgSerAlaIleCysArg	816
2627	GTCTGTGACTGTGGTGTGGATCCAGAGNAGAAACAGGTGTGTCAA	2676
817	LysMetLeuLysThrGlyLeuSerThrValValAsnSerThrLeuCysPr	833
2677	AGCTGGCAGCAAAAGCTCGCGCATCCCTCTCAGTGAATGATGTGCAG	2726

4378AATGACCCA..... 4386
1419 rotleysGlyHisProValProAsnIleThrTrpPheHisGlyGlyGln 1435
4387ACAGGAGAACCCCGCCCT..... 4404
1436 ProileValThrAlaThrGlyLeuThrHisHisIleLeuAlaAlaGlyG1 1452
4404 4404
1452 nIleLeuGlnValAlaAsnLeuSerGlyGlySerGlnGlyGluPheSerC 1469
4404 4404
1469 ysLeuAlaGlnAsnGluAlaGlyValLeuMetGlnLysAlaSerLeuVal 1485
4404 4404
1486 IleClnAspTyrTrpTrpSerValAspArgLeuAlaThrCysSerAlaSe 1502
4405 ...CAAGAGCCTTTGGGAGCCTGGTAACTGGTGCATCTGTTCTGCGCAC 4451
1502 rCysGlyAsnArgGlyValGlnGlnProArgLeuArgCysLeuLeu...A 1518
4452 CTGTGGTCATTGGGAGCCCGCATTCAGAGACCCCGAGTGTGTATGGCCA 4501
1518 snSerThrGluValAsnProAlaHisCysAlaGlyLysValArgProAla 1534
4502 ATGGGAGGAGGAGTGAGTGGCCCTGTGTGATCACCITCCAGAGCCACTG 4551
1535 ValGlnProIleAlaCysAsnArgArgAspCysProSerArgTrpMetVa 1551
4552 GCTGGGTTTGGCCCTGTAACATCCGGGACTGCCAGGAGGTGGTTCAC 4601
1551 lThrSerTrpSerAlaCysThrArgSerCysGlyGlyGlyValGlnThrA 1568
4602 AAGTGTGTGTGTGTCAGGTCTGTGTCTTTCGGGTGAAGGATACACAGTC 4651
1568 rArgValThrCysGlnLysLeuLysAlaSerGlyIleSerThrProVal 1584
4652 GGCAGGTGACGTCAACCGCACAAACCAATGGAAGTGTGCAGGTGGTG 4701
1585 SerAsnAspMetCysThrGlnValAlaLysArgProValAspThrGlnAl 1601
4702 TCTCCAAGAGCATGTGCCCT...AAGACCGGCTCTGGGAAGAAACCC 4748
1601 aCysAsnGlnClnLeuCysValGluTrpAlaPheSerSerTrpGlyGlnC 1618
4749 ATGTTTGGTCAATCCATGTGTGTAGTGG.....GAACGAGGAAACCGT 4792
1618 ysAsnGlyProCysIleGlyProHisLeuAlaValGlnHisArgGlnVal 1634
4793 GTCTGGACGTTGATGGCGGTGCTGTGAGGATGCAGCAGCGTCACACA 4842
1635 PheCysGlnThrArgAspGlyIleThrLeuProSerGluGlnCysSerAl 1651
4843 GCTTGTCAACACACAGCTCT.....GACTCCAACCTGTGATGA 4880
1651 aLeuProArgProValSerThrGlnAsnCysTrpSerGluAlaCysSerV 1668
4881 CAGAAAGAGACCCACCTTAAGAGGAAGTGCATCATCAGGGCCCTGTGATG 4930
1668 alHisTrpArgValSerLeuTrpThrLeuCysThrAlaThrCysGlyAsn 1684
4931 TGTGTGSCACACAGCGCTTGAAGCCCTGTACAGCAGCCTGTGGCAGG 4980
1685 TyrGlyPheGlnSerArgArgValGluCysValHisAlaArgThrAsnLy 1701
4981 ...GGTTTCCAGTCTCGGAAAGTCACTGTATCCACAAAGAGTGGCA 5027
1701 sAlaValProGluLeuLeuCysSerTrpGlyProArgProAlaAsnTrpG 1718
5028 ACCTGTGCCAAGAGACACTGTGTACAGAAAGAAAGAACCAATTCCTGGC 5077

1718 lnArgCysAsnIleThrProCysGluAsnMetGluCysArgAspThrThr 1734
5078 GGCACCT...CTTGGGCCCTCCTGTGATAGACTGCACAGACACAACT 5124
1735 ArgTyrCysGluLysValLysGlnLeuLysLeuCysGlnLeuSerGlnPh 1751
5125 CACTACTGTATGTTTGAACATCTTAATTTGTTCTCTAGACCGCTA 5174
1751 eLysSerArgCysCysGlyThrCys 1759
5175 CAAACAAGGTGTCGCCAGTCATGT 5199

seq_name: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA2001A.DAT:AAD14371

seq_documentation_block:

ID AAD14371 standard; cDNA; 4854 BP.

XX AAD14371;

XX DT 01-NOV-2001 (first entry)

XX DE Novel human protein (NHP) cDNA #8.

XX KW Novel human protein; NHP; thrombospondin; gene therapy; cancer;
KW cytostatic; antisense therapy; angiogenesis; biological disorder; ss.
XX OS Homo sapiens.

XX FH Key Location/Qualifiers

XX CDS 1..4854

XX FT /*tag= a

XX FT /product= "Novel human protein"

XX PN WO200161011-A2.

XX PD 23-AUG-2001.

XX PF 15-FEB-2001; 2001WO-US05290.

XX PR 17-FEB-2000; 2000US-0183282.

XX PA (LEXI-) LEXICON GENETICS INC.

XX PI Donoho G, Scoville J, Turner CA, Friedrich G, Zambrowicz B;
PI Sands AT;

XX DR WPI; 2001-514776/56.

XX DR P-PSDB; AAE07870.

XX PT An isolated nucleic acid encoding a thrombospondin useful as a
PT hybridization probe and gene therapy treatments of cancer -

XX PS Claim 4; Page 48-49; 56pp; English.

XX CC The invention relates to novel human polynucleotides encoding proteins
CC that share sequence similarity with animal proteins having thrombospondin
CC repeats. NHPs are expressed in human cell lines, pituitary, lymph nodes,
CC prostate, testis, adrenal gland, uterus, foetal kidney, foetal lung and
CC gene trapped human cells. NHPs are thrombospondins useful for treating
CC biological disorders involving angiogenesis, cancer and development and
CC also in pharmacogenomic applications. NHPs are useful as a hybridisation
CC probe for screening libraries, assessing gene expression patterns and
CC also in gene therapy. Proteins having thrombospondin repeats act as
CC receptors, secreted extracellular matrix proteins and proteases. The
CC present sequence is a cDNA encoding novel human protein (NHP).

XX SQ Sequence 4854 BP; 1331 A; 1175 C; 1276 G; 1072 T; 0 other;

alignment_scores:

Quality: 3762.50

Ratio: 3.194

Length: 1733

Gaps: 31

Percent Similarity: 67.975 Percent Identity: 42.527

alignment_block:

US-10-044-807-2 x AAD14371 ..

Align seg 1/1 to: AAD14371 from: 1 to: 4854

```
64  LysSerCysGluGlyArgAsnIleArgTyrArgThrCysSerAsnValAs 80
      ::::::::::::::::::::::::::::::::::::::::::::::::::::
94  AGGAATTTGTAAGGGCAGACAATCGGTACAGACATGCAGCAATCATGA 143
      ::::::::::::::::::::::::::::::::::::::::::::::::::::::
80  pCysProGluAlaGlyAspPheArgAlaGlnGlnCysSerAlaHisA 97
      ::::::::::::::::::::::::::::::::::::::::::::::::::::::
144 CTGCCCTCCAGATGCAGAGATTTCAGAGCCACAGCTGCTCAGCCCTACA 193
      ::::::::::::::::::::::::::::::::::::::::::::::::::::::
97  snAspValLysHisHisGlyGlnPheTyrGluTrpLeuProValSerAsn 113
      ::::::::::::::::::::::::::::::::::::::::::::::::::::::
194 ATGATGTCCAGTATCAGGGCATTACTATGAATGGCTTCCACGATATAAT 243
      ::::::::::::::::::::::::::::::::::::::::::::::::::::::
114 AspProAspAsnProCysSerLeuLysCysGlnAlaLysGlyThrLe 130
      ::::::::::::::::::::::::::::::::::::::::::::::::::::::
244 GATCCCTGCTGCCCGGTGTCACCTCAAGTGTGATGGAACCTGTCGAACACGG 343
      ::::::::::::::::::::::::::::::::::::::::::::::::::::::
130 uValValGluLeuAlaProLysValLeuAspGlyThrArgCysTyrThrG 147
      ::::::::::::::::::::::::::::::::::::::::::::::::::::::
294 GGTGGTGGAGCTGGCACCCTAAGGTACTGGATGGAACCTGTCGAACACGG 343
      ::::::::::::::::::::::::::::::::::::::::::::::::::::::
147 luSerLeuAspMetCysIleSerGlyLeuCysGlnIleValGlyCysAsp 163
      ::::::::::::::::::::::::::::::::::::::::::::::::::::::
344 ACTCCTTGGACATGTGTATCAGTGGCATCTGTCAGGCAGTGGGCTCGAT 393
      ::::::::::::::::::::::::::::::::::::::::::::::::::::::
164 HisGlnLeuGlySerThrValLysGluAspAsnCysGlyValCysAsnG 180
      ::::::::::::::::::::::::::::::::::::::::::::::::::::::
394 CGCAACTGGGAAGCACTGCAAGAGGAGCAACTGTGGAGTCTGTCCGG 443
      ::::::::::::::::::::::::::::::::::::::::::::::::::::::
180 yAspGlySerThrCysArgLeuValArgGlyGlnTyrLysSerGlnLeu 197
      ::::::::::::::::::::::::::::::::::::::::::::::::::::::
444 CGATGCTCCAGCTCAGGCTGTGCGGGACAACTCAAGATCAGCGTTT 493
      ::::::::::::::::::::::::::::::::::::::::::::::::::::::
197 erAlaThrLysSerAspThrValValAlaIleProTyrGlySerArg 213
      ::::::::::::::::::::::::::::::::::::::::::::::::::::::
494 CTCCTCAAAAAGAGAGAATAATGTAATGCTGTCTCTTTGGGAAGTCGA 543
      ::::::::::::::::::::::::::::::::::::::::::::::::::::::
214 HisIleArgLeuValLysGlyProAspHisLeuTyrLeuGluThrLy 230
      ::::::::::::::::::::::::::::::::::::::::::::::::::::::
544 AGTGTGAGAAATTACAGTGAAGAGCACTGCCACCTCTTTATTCGAATCAA 593
      ::::::::::::::::::::::::::::::::::::::::::::::::::::::
230 sThrLeuGlnGlyThrLysGlyGluAsnSerLeuSerThrGlyThrP 247
      ::::::::::::::::::::::::::::::::::::::::::::::::::::::
594 AACACTTCAGGAGCAAGAGGAGCAACACAGCTTTAACAGCCCGGGGTCT 643
      ::::::::::::::::::::::::::::::::::::::::::::::::::::::
247 heLeuValAspAsnSerSerValAspPheGlnLysPheProAspLysGlu 263
      ::::::::::::::::::::::::::::::::::::::::::::::::::::::
644 TTGCTGTAGAAAACACACAGTGAATTTTCAGAGGGGCTCCGAGAGGCCAA 693
      ::::::::::::::::::::::::::::::::::::::::::::::::::::::
264 IleLeuArgMetAlaGlyProLeuThrAlaAspPheIleValLysIle 280
      ::::::::::::::::::::::::::::::::::::::::::::::::::::::
694 ACTTTTAAGATTCCAGGAGCTCTGTATGCTGATTTTCATCTTCAAGACCAG 743
      ::::::::::::::::::::::::::::::::::::::::::::::::::::::
280 gAsnSerGlySerAlaAspSerThrValGlnPheIlePheTyrGlnProI 297
      ::::::::::::::::::::::::::::::::::::::::::::::::::::::
744 GTACACTGCAGCCCAAGACAGCGTGTTCAGTTCCTTTTACAGCCCA 793
      ::::::::::::::::::::::::::::::::::::::::::::::::::::::
297 leIleHisArgTrpArgGluThrAspPhePheProCysSerAlaThrCys 313
      ::::::::::::::::::::::::::::::::::::::::::::::::::::::
794 TCAGTTCATCAGTGGACACAACTGACTTCTTCCCTGCACTGTGACGTGT 843
      ::::::::::::::::::::::::::::::::::::::::::::::::::::::
314 GlyGlyGlyTyrGlnLeuThrSerAlaGluCysTyrAspLeuArgSerAs 330
      ::::::::::::::::::::::::::::::::::::::::::::::::::::::
844 GGAGGAGGTTATCAGTCAATTTCTGCTGAATGTGTGATATCCGCTTGAA 893
      ::::::::::::::::::::::::::::::::::::::::::::::::::::::
330 nArgValValAlaAspGluTyrCysHisTyrTyrProGluAsnIleLysP 347
      ::::::::::::::::::::::::::::::::::::::::::::::::::::::
```

```
894 GAGGTAGTTCTCTGACCATATTGTCTACTACTACCTGMAAATGTA AAC 943
347 roLysProLysLeuGlnGluCysAsnLeuAspProCysProAlaSerAsp 363
      ::::::::::::::::::::::::::::::::::::::::::::::::::::::
944 CAAACCAAACTGAAGGAATGCAGCATGGATCCCTCCCATCAAGTCAT 993
      ::::::::::::::::::::::::::::::::::::::::::::::::::::::
364 GlyTyrLysGlnIleMetProTyrAspLeuTyrHisProLeuProArgTr 380
      ::::::::::::::::::::::::::::::::::::::::::::::::::::::
994 GGATTTAAAGAGATAATGCCCTATGACCACTTCCAACTCTTCTCTGCTG 1043
      ::::::::::::::::::::::::::::::::::::::::::::::::::::::
380 pGluAlaThrProTrpThrAlaCysSerSerCysGlyGlyGlyLeG 397
      ::::::::::::::::::::::::::::::::::::::::::::::::::::::
1044 GGAACATAATCTTGGACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1093
      ::::::::::::::::::::::::::::::::::::::::::::::::::::::
397 lnSerArgAlaValSerCysValGluGluAspIleGlnGlyHisValThr 413
      ::::::::::::::::::::::::::::::::::::::::::::::::::::::
1094 AGAGACGGAGCTTGTGTGTGTAGAGAAATCCATGTCATGGAGAGATATTG 1143
      ::::::::::::::::::::::::::::::::::::::::::::::::::::::
414 SerValGluGluTrpLysCysMetTyrThrProLysMetProIleAlaG 430
      ::::::::::::::::::::::::::::::::::::::::::::::::::::::
1144 CAGGTGCAAGAATGGAAGTGCATGTACGCACCAACCAAGCTTATGCA 1193
      ::::::::::::::::::::::::::::::::::::::::::::::::::::::
430 nProCysAsnIlePheAspCysProLysTrpLeuAlaGlnGluTrpSerP 447
      ::::::::::::::::::::::::::::::::::::::::::::::::::::::
1194 AACTTGTAACTCTGTTGATTCGCCCAAGTGGATTGCCATGGAGTGTCTC 1243
      ::::::::::::::::::::::::::::::::::::::::::::::::::::::
447 roCysThrValThrCysGlyGlnGlyLeuArgTyrArgValValLeuCys 463
      ::::::::::::::::::::::::::::::::::::::::::::::::::::::
1244 AGTGACAGTGCATGTGGCCGAGGGTTACGGTACCGGGTTGTCTGTGT 1293
      ::::::::::::::::::::::::::::::::::::::::::::::::::::::
464 IleAspHisArgGlyMetHisThrGlyGlyCysSerProLysThrLysPr 480
      ::::::::::::::::::::::::::::::::::::::::::::::::::::::
1294 ATTAACCAACCGCGAGAGCATGTGTGGGGCTGCAATCCCAACTGAAGTT 1343
      ::::::::::::::::::::::::::::::::::::::::::::::::::::::
480 oHisIleGluGluCysIleValProThrProCysTyrLysProLysG 497
      ::::::::::::::::::::::::::::::::::::::::::::::::::::::
1344 ACATCTCAAAAGAAGATGTGTCTATCCATCCCGTGTATATAACCAAAAG 1393
      ::::::::::::::::::::::::::::::::::::::::::::::::::::::
497 luLysLeuProValGluAlaLysLeuProThrPheLysGlnAlaGlnGlu 513
      ::::::::::::::::::::::::::::::::::::::::::::::::::::::
1394 AAAAAAGTCCAGTGGAAAGCAAAATTCCTTGGCTGAAACCAAGCAAGAA 1443
      ::::::::::::::::::::::::::::::::::::::::::::::::::::::
514 LeuGluGluGlyAlaAlaValSerGluGluProSerPheIleProGluAl 530
      ::::::::::::::::::::::::::::::::::::::::::::::::::::::
1444 CTAGAAGAGACCAAGAAATAGCAACAGAGAACCACAGCTTCATTCAGAAC 1493
      ::::::::::::::::::::::::::::::::::::::::::::::::::::::
530 atTpSerAlaCysThrValThrCysGlyValGlyThrGlnValArgIleV 547
      ::::::::::::::::::::::::::::::::::::::::::::::::::::::
1494 CTGGTCAGCCTGCAGTACCACGTGTGGGCCAGGTGTGCAGGTCCGCGAG 1543
      ::::::::::::::::::::::::::::::::::::::::::::::::::::::
547 alArgCysGlnValLeuLeuSerPheSerGlnSerValAlaAlaAspLeuPro 563
      ::::::::::::::::::::::::::::::::::::::::::::::::::::::
1544 TGAAGTGCCTGTGCTCTCATATTCAGCAGACTTGAGACTGAGCTGAGCTGCC 1593
      ::::::::::::::::::::::::::::::::::::::::::::::::::::::
564 IleAspGluCysGluGlyProLysProAlaSerGlnArg..... 576
      ::::::::::::::::::::::::::::::::::::::::::::::::::::::
1594 GAGGAAGAGTGTGAAGGCCCAAGCTGCCCAAGCAAGCGGCTGCCTCTCT 1643
      ::::::::::::::::::::::::::::::::::::::::::::::::::::::
577 .....AlaCysTyrAlaGlyProCysSerGlyGluIleProGluPheAsn 592
      ::::::::::::::::::::::::::::::::::::::::::::::::::::::
1644 GGAAGCATGTGATGAGAGCCCGGCTCCCGAGAGACTAGACATCCCTCTCC 1693
      ::::::::::::::::::::::::::::::::::::::::::::::::::::::
592 roAspGluThrAspGlyLeuPheGlyLeuGlnAspPheAspGluLeu 608
      ::::::::::::::::::::::::::::::::::::::::::::::::::::::
1694 CT.....GAGGACAGTGAAGCACT 1713
      ::::::::::::::::::::::::::::::::::::::::::::::::::::::
609 TyrAspTrpGluTyrGluGlyPheThrLysCysSerGlySerCysGlyGl 625
      ::::::::::::::::::::::::::::::::::::::::::::::::::::::
1714 TACGACTGGGAGTACCTGGGTTTACCCCTTGACACCAACATGCTTGGG 1763
      ::::::::::::::::::::::::::::::::::::::::::::::::::::::
625 yGlyValGlnGluAlaValSerCysLeuAsnLysGlnThrArgGluP 642
      ::::::::::::::::::::::::::::::::::::::::::::::::::::::
1764 AGGCCATCAAGAAGCCATAGCAGTGTGCTTACATATCCAGACCCAGCAGA 1813
      ::::::::::::::::::::::::::::::::::::::::::::::::::::::
```

[illegible]

```

3464 AGNATGAACTTGTTCAGCCCTCAGTAAATAATTTGGATGGAACT 3513
      :::::||||| :::::||||| :::::||||| :::::||||| :::::|||||
1233 AspSerLeuGlnIleLeuAlaProValGluAlaAspValGlyPheTyrTh 1249
      :::::||||| :::::||||| :::::||||| :::::||||| :::::|||||
3514 GGGAAAGATACAGATACAGATCTCACAAGGAAAGAACAGGCATATATGA 3563
      :::::||||| :::::||||| :::::||||| :::::||||| :::::|||||
1249 rCysAsnAlaThrAsnAlaLeuGlyTyrAspSerValSerIleAlaValT 1266
      :::::||||| :::::||||| :::::||||| :::::||||| :::::|||||
3564 ATGTTCTGTAGCTAATCATCTGTGGTTCAGATGTGGAAGTCTTCTGTGC 3613
      :::::||||| :::::||||| :::::||||| :::::||||| :::::|||||
1266 hrLeuAlaGlyLysProLeuValLysThrSerArgMetThrValIleAsn 1282
      :::::||||| :::::||||| :::::||||| :::::||||| :::::|||||
3614 TGTATGAGAGGCACCTGTCTATCTGTCTGTGTAAGAAATATACCCAAA 3663
      :::::||||| :::::||||| :::::||||| :::::||||| :::::|||||
1283 ThrGluLysProAlaValThrValAspIleGlySerThrIleLysThrVa 1299
      :::::||||| :::::||||| :::::||||| :::::||||| :::::|||||
3664 CCAGGCACAAACCATCTGTCTGTGTGGTGGAGGCATCTGGAGGCAGC 3713
      :::::||||| :::::||||| :::::||||| :::::||||| :::::|||||
1299 lGlnGlyValAsnValThrIleAsnCysGlnValAlaGlyValProGluA 1316
      :::::||||| :::::||||| :::::||||| :::::||||| :::::|||||
3714 CMTGGAGCAACGTCACATCCGATCTCTGTGTAAGAGGTGTCCTCAGC 3763
      :::::||||| :::::||||| :::::||||| :::::||||| :::::|||||
1316 laGluValThrTrpPheArgAsnLysSerLysLeu...GlySerProHis 1331
      :::::||||| :::::||||| :::::||||| :::::||||| :::::|||||
3764 CTAATATACTTGTGTAAGAGGAGGATCTCTGAGTGGCAATGTTTCC 3813
      :::::||||| :::::||||| :::::||||| :::::||||| :::::|||||
1332 HisLeuHisGluGlySerLeuLeuLeuThrAsnValSerSerAspG1 1348
      :::::||||| :::::||||| :::::||||| :::::||||| :::::|||||
3814 TTGCTTTTCAATGGATCCCTGTTGTCAGATGTTTCCCTTGAANATGA 3863
      :::::||||| :::::||||| :::::||||| :::::||||| :::::|||||
1348 nGlyLeuTyrSerCysArgAlaAlaAsnLeuHisGlyGluLeuThrGluS 1365
      :::::||||| :::::||||| :::::||||| :::::||||| :::::|||||
3864 AGGAACCTAGCTGTGATAGCAACCAATGCTCTTGGAAAGCGATGGCAA 3913
      :::::||||| :::::||||| :::::||||| :::::||||| :::::|||||
1365 erThrGlnLeuLeuIleLeuAspProProGlnValProThrGlnLeuGlu 1381
      :::::||||| :::::||||| :::::||||| :::::||||| :::::|||||
3914 CATCTGACTCCACTTCTGGAGAGAGATGGCCAGAGATAGTAATCGTA 3963
      :::::||||| :::::||||| :::::||||| :::::||||| :::::|||||
1382 AspIleArgAla.....LeuLeuAlaAlaThrGlyProAsnLe 1394
      :::::||||| :::::||||| :::::||||| :::::||||| :::::|||||
3964 TTTCTGCAAGGACATAAAAGTAGTATCTCTCCAGGCAACCACTAGAAC 4013
      :::::||||| :::::||||| :::::||||| :::::||||| :::::|||||
1394 uProSerValLeuThrSerProLeuGlyThrGlnLeuValLeuAspProG 1411
      :::::||||| :::::||||| :::::||||| :::::||||| :::::|||||
4014 CAACAC.....AATGACCCCA. 4029
      :::::||||| :::::||||| :::::||||| :::::||||| :::::|||||
1411 lYasnSerAlaLeuLeuGlyCysProIleLysGlyHisProValProAsn 1427
      :::::||||| :::::||||| :::::||||| :::::||||| :::::|||||
4030 .....ACAGGAGAACCCCGCCT... 4047
      :::::||||| :::::||||| :::::||||| :::::||||| :::::|||||
1428 lIleThrTrpPheHisGlyGlyGlnProIleValThrAlaThrGlyLeuTh 1444
      :::::||||| :::::||||| :::::||||| :::::||||| :::::|||||
4047 ..... 407
      :::::||||| :::::||||| :::::||||| :::::||||| :::::|||||
1444 rHisHisIleLeuAlaAlaGlyGlnIleLeuGlnValAlaAsnLeuSerG 1461
      :::::||||| :::::||||| :::::||||| :::::||||| :::::|||||
4047 ..... 407
      :::::||||| :::::||||| :::::||||| :::::||||| :::::|||||
1461 lYgGlySerGlnGlyGluPheSerCysLeuAlaGlnAsnGluAlaGlyVal 1477
      :::::||||| :::::||||| :::::||||| :::::||||| :::::|||||
4047 ..... 407
      :::::||||| :::::||||| :::::||||| :::::||||| :::::|||||
1478 LeuMetGlnLysAlaSerLeuValIleGlnAspTyrTrpTrpSerValAs 1494
      :::::||||| :::::||||| :::::||||| :::::||||| :::::|||||
4048 .....CAAGAGCCCTTTTGGAGCGCTGG 4070
      :::::||||| :::::||||| :::::||||| :::::||||| :::::|||||
1494 pArgLeuAlaThrCysSerAlaSerCysGlyAsnArgGlyValGlnGlnP 1511
      :::::||||| :::::||||| :::::||||| :::::||||| :::::|||||
4071 TAACTGTGCACATGTGTGCGCACCTGTGTCATTTGGAGCGCCGCAATC 4120
      :::::||||| :::::||||| :::::||||| :::::||||| :::::|||||
1511 roArgLeuArgCysLeuLeu...AsnSerThrGluValAsnProAlaHis 1526
      :::::||||| :::::||||| :::::||||| :::::||||| :::::|||||

```

```

4121 AGAGACCCCACTGTGTGATGGCCAATGGCAGGAAGTGAGTGAGCCCTG 4170
      :::::||||| :::::||||| :::::||||| :::::||||| :::::|||||
1527 CysAlaGlyLysValArgProAlaValGlnProIleAlaCysAsnArgAr 1543
      :::::||||| :::::||||| :::::||||| :::::||||| :::::|||||
4171 TGTGATCACCTCCAGAACCCACTGGCTGGTGGTTGAGCCCTGTAAACATCG 4220
      :::::||||| :::::||||| :::::||||| :::::||||| :::::|||||
1543 gAspCysProSerArgTrpMetValThrSerTrpSerAlaCysThrArgS 1560
      :::::||||| :::::||||| :::::||||| :::::||||| :::::|||||
4221 GCACCTGCCACCGAGGTGGTTTCAACAAGTGTGGTTCACAGTCTCTGTGT 4270
      :::::||||| :::::||||| :::::||||| :::::||||| :::::|||||
1560 erCysGlyGlyGlyValGlnThrArgArgValThrCysGlnLysLeuLys 1576
      :::::||||| :::::||||| :::::||||| :::::||||| :::::|||||
4271 CTTCCGGTGAAGGATACACAGTCGCGAGGTGACGTGCAAGCGGACAAA 4320
      :::::||||| :::::||||| :::::||||| :::::||||| :::::|||||
1577 AlaSerGlyIleSerThrProValSerAsnAspMetCysThrGlnValAl 1593
      :::::||||| :::::||||| :::::||||| :::::||||| :::::|||||
4321 GCCAATGGAACCTGTCAGGTGTGTCTCCAAGAGCATGTGCCCT...AA 4367
      :::::||||| :::::||||| :::::||||| :::::||||| :::::|||||
1593 aLysArgProValAspThrGlnAlaCysAsnGlnGlnLeuCysValGluT 1610
      :::::||||| :::::||||| :::::||||| :::::||||| :::::|||||
4368 AGACCGCCCTCTGGGAAGAAACCATGTTTGGTTCATCCATGTTTCACT 4417
      :::::||||| :::::||||| :::::||||| :::::||||| :::::|||||
1610 rPalapheSerSerTrpGlyGlnCysAsnGlyProCysIleGlyProHis 1626
      :::::||||| :::::||||| :::::||||| :::::||||| :::::|||||
4418 GG.....GAACCCAGGGAACCGGTCTCTGACGTTGCATGGCGGTGCT 4461
      :::::||||| :::::||||| :::::||||| :::::||||| :::::|||||
1627 LeuAlaValGlnHisArgGlnValPheCysGlnThrArgAspGlyIleTh 1643
      :::::||||| :::::||||| :::::||||| :::::||||| :::::|||||
4462 GTGAGGATGACGACGCGTCACACAGCTTGTCAACACACACAGCTCT... 4506
      :::::||||| :::::||||| :::::||||| :::::||||| :::::|||||
1643 rLeuProSerGluGlnCysSerAlaLeuProArgProValSerThrGlnA 1660
      :::::||||| :::::||||| :::::||||| :::::||||| :::::|||||
4507 .....GACTCCAACCTGTGATGACAGAAAGACACCCACCTTAAGAGGA 4549
      :::::||||| :::::||||| :::::||||| :::::||||| :::::|||||
1660 snCysTrpSerGluAlaCysSerValHisTrpArgValSerLeuTrpThr 1676
      :::::||||| :::::||||| :::::||||| :::::||||| :::::|||||
4550 ACTGCACATCAGGGGCTGTGATGTGTGTTGGCACACAGGCGCTTGAAG 4599
      :::::||||| :::::||||| :::::||||| :::::||||| :::::|||||
1677 LeuCysThrAlaThrCysGlyAsnTyrGlyPheGlnSerArgArgValG 1693
      :::::||||| :::::||||| :::::||||| :::::||||| :::::|||||
4600 CCCTGTACACAGCGCTGTGCAGG...GGTTTCCAGTCTCGGAAAGTCGA 4646
      :::::||||| :::::||||| :::::||||| :::::||||| :::::|||||
1693 uCysValHisAlaAargThrAsnLysAlaValProGluHisLeuCysSerT 1710
      :::::||||| :::::||||| :::::||||| :::::||||| :::::|||||
4647 CTGTATCCACACAAAGAGTGTGCAACCTGTGGCCACACAGACACTGTGTAC 4696
      :::::||||| :::::||||| :::::||||| :::::||||| :::::|||||
1710 rpGlyProArgProAlaAsnTrpGlnArgCysAsnIleThrProCysGlu 1726
      :::::||||| :::::||||| :::::||||| :::::||||| :::::|||||
4697 AGAAAAGAAACCAATTTCTGTGGCGCACTGT...CTTGGGCGCTCCTGT 4743
      :::::||||| :::::||||| :::::||||| :::::||||| :::::|||||
1727 AsnMetGluCysArgAspThrThrArgTyrCysGluLysValLysGlnLe 1743
      :::::||||| :::::||||| :::::||||| :::::||||| :::::|||||
4744 GATAGAGACTGCACAGACACAACTCACTACTGTATGTTGTAAACATCT 4793
      :::::||||| :::::||||| :::::||||| :::::||||| :::::|||||
1743 uLysLeuCysGlnLeuSerGlnPheLysSerArgCysCysGlyThrCys 1759
      :::::||||| :::::||||| :::::||||| :::::||||| :::::|||||
4794 TAATTGTGTCTCTAGACCGCTACAAACAAAGGTGCTGCGCAGCTCATGT 4842
      :::::||||| :::::||||| :::::||||| :::::||||| :::::|||||

```

seq_name: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA2001A.DAT:AAD14372

seq_documentation_block:

ID AAD14372 standard; DNA; 8578 BP.

AC AAD14372;

DT 01-NOV-2001 (first entry)

XX Novel human protein gene open reading frame with flanking sequence.

DE Novel human protein; NHP; thrombospondin; gene therapy; cancer;

KW cytostatic; antisense therapy; angiogenesis; biological disorder; ds.

XX Homo sapiens.

OS

XX WO200161011-A2.
 PN 23-AUG-2001.
 PD 15-FEB-2001; 2001WO-US05290.
 PF 17-FEB-2000; 2000US-0183282.
 PR (LEXI-) LEXICON GENETICS INC.
 XX
 XX Donoho G, Scoville J, Turner CA, Friedrich G, Zambrowicz B;
 PI Sands AT;
 XX
 XX WPI; 2001-514776/56.
 DR
 XX An isolated nucleic acid encoding a thrombospondin useful as a
 PI hybridization probe and gene therapy treatments of cancer -
 XX
 XX Disclosure; Page 53-56; 56pp; English.
 XX
 CC The invention relates to novel human polynucleotides encoding proteins
 CC that share sequence similarity with animal proteins having thrombospondin
 CC repeats. NHPs are expressed in human cell lines, pituitary, lymph nodes,
 CC prostate, testis, adrenal gland, uterus, foetal kidney, foetal lung and
 CC gene trapped human cells. NHPs are thrombospondins useful for treating
 CC biological disorders involving angiogenesis, cancer and development and
 CC also in pharmacogenomic applications. NHPs are useful as a hybridisation
 CC probe for screening libraries, assessing gene expression patterns and
 CC also in gene therapy. Proteins having thrombospondin repeats act as
 CC receptors, secreted extracellular matrix proteins and proteases. The
 CC present sequence is novel human protein (NHP) gene containing NHP
 CC open reading frame with flanking sequences.
 XX
 XX Sequence 8578 BP; 2440 A; 1827 C; 2070 G; 2241 T; 0 other;

alignment_scores:
 Quality: 3673.00 Length: 2207
 Ratio: 3.016 Gaps: 36
 Percent Similarity: 55.188 Percent Identity: 34.798

alignment_block:
 US-10-044-807-2 x AAD14372 ..

Align seg 1/1 to: AAD14372 from: 1 to: 8578

22 SerSerArgThrAlaArgSerGluGluAspArgAspGlyLeuTrpAspAl 38
 ::::|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
 368 ACCTCAAGAACACTCGTTTCAGATGAAGACAAAGATGGCAACTGGGATGC 417
 ||||| |||||:|||||:|||||:|||||:|||||:|||||:|||||
 38 aTrpGlyProTrpSerGluCysSerArgThrCysGlyGlyAlaSerT 55
 ||||| |||||:|||||:|||||:|||||:|||||:|||||:|||||
 418 TTGGGGCGACGAGAGTACGCTCCCGGACCTGTGGGGGAGGACATCAT 467
 ||||| |||||:|||||:|||||:|||||:|||||:|||||:|||||
 55 YrSerLeuArgCysCysLeuSerSerLysSerCysGluGlyArgAsnIle 71
 :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
 468 ATTCTCTCGGAGATGTTTGCATGGAAGGATTTGTGAAGGGCAGACATT 517
 ||||| |||||:|||||:|||||:|||||:|||||:|||||:|||||
 72 ArgTrpArgThrCysSerAsnValAspCysProProGluAlaGlyAspPh 88
 :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
 518 CGGTACAGAGATCGACATCATGCTGCTCCCTCCAGATGAGAGATTT 567
 ||||| |||||:|||||:|||||:|||||:|||||:|||||:|||||
 88 eArgAlaGlnCysSerAlaHisAsnAspValLysHisGlyGlnP 105
 :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
 568 CAGAGCCAGCAGTGTCTCAGCTACATGATGTCCTCAGTATCAGGGCATT 617
 ||||| |||||:|||||:|||||:|||||:|||||:|||||:|||||
 105 heTrpGluTrpLeuProValSerAsnAspProAspAsnProCysSerLeu 121
 :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
 618 ACTATGAATGGCTCCACGATATATATGCTGCTGCCCGGTGTGCATC 667
 ||||| |||||:|||||:|||||:|||||:|||||:|||||:|||||
 122 LysCysGlnAlaLysGlyThrThrLeuValValGluLeuAlaProLysVa 138
 :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||

668 AAGTGTCTATCCACAGACAAAACCTTGGTGGTGGAGCTGCACCTAAGGT 717
 138 lLeuAspGlyThrArgCysTyrThrGluSerLeuAspMetCysIleSerG 155
 :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
 718 ACTGGATGGAACCTCGTTGGCAACACGGGACTCCTTGGACATGTGTATCAGTG 767
 155 lyLeuCysGlnIleValGlyCysAspHisGlnLeuGlySerThrValLys 171
 :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
 768 GCATCTGTACGGCAGTGGCTGCGATCGCACTGGGAACATGCCAAG 817
 172 GluAspAsnCysGlyValCysAsnGlyAspGlySerThrCysArgLeuVa 188
 :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
 818 GAGACAACTGTGGAGTCTGTGCGCGCATGGCTCCACCTGCAGGCTTGT 867
 188 lArgGlyGlnTyrLysSerGlnLeuSerAlaThrLysSerAspAspThrV 205
 :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
 868 ACGGGGACAACTCAAGTCAACAGCTTCTCTGAAAAAAGAGAGAGAAAATG 917
 205 alValAlaIleProTyrGlySerArgHisIleArgLeuValLeuLysGly 221
 :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
 918 TAATTGTCTTCTTCTTGGGAAGTCGAAGTGTGAGATTACAGTGAAGA 967
 222 ProAspHisLeuTyrLeuGluThrLysThrLeuGlnGlyThrLysGlyG 238
 :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
 968 CTTCCCCACCTCTTATTGATCAAAAACACTTCAAGGAACCAAGGAGA 1017
 238 uAsnSerLeuSerSerThrGlyThrPheLeuValAspAsnSerSerVala 255
 :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
 1018 ACACAGCTTTAACAGCCCGCGCTTTTGTGCTAGAAAACACAAACAGTGG 1067
 255 spPheGlnLysPheProAspLysGluIleLeuArgMetAlaGlyProLeu 271
 :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
 1068 AATTTCAGAGGGCTCCGAGAGAGCAAACTTTTAAGATTCCAGGACCTCTG 1117
 272 ThrAlaAspPheIleValLysIleArgAsnSerGlySerAlaAspSerTh 288
 :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
 1118 ATGGCTGATTTCATCTTCAACACCGGTACACTGCAGCCCAAGACAGCGT 1167
 288 rValGlnPheIlePheTyrGlnProIleIleHisArgTrpArgGluThrA 305
 :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
 1168 GGTTCAGTTCTTCTTTTACCAGCCCATCAGTCATCAGTGGAGACAACTG 1217
 305 spPhePheProCysSerAlaThrCysGlyGlyGlyTyrGlnLeuThrSer 321
 :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
 1218 ACTTCTTCCCTGCATGTGAGTGTGGAGGAGTTATCAGCTCAATTTCT 1267
 322 AlaGluCysTyrAspLeuArgSerAsnArgValValAlaAspGlnTyrCy 338
 :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
 1268 CTTAAATGTGTGGATATCCGCTTGAGAGAGGTAGTTCCTGACCATATTG 1317
 338 sHisTyrTyrProGluAsnIleLysProLysProLysLeuGlnCysA 355
 :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
 1318 TCACCTACTACCTGAAAAATGTAAACCAAAACCAAACTGAAGGAATGCA 1367
 355 snLeuAspProCysProAlaSerAspGlyTyrLysGlnIleMetProTyr 371
 :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
 1368 GCATGGATCCCTGCCCATCAAGTGAATGATTTAAAGAGATTAATGCCCTAT 1417
 372 AspLeuTyrHisProLeuProArg..TrpGluAlaThrProTyrTrpAla 387
 :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
 1418 GACCACCTTCCAAACCTTCTTCTCGAGCTGGAAACATAATCCTTGGACTGCA 1467
 388 CysSerSerCysGlyGlyGlyIleGlnSerArgAlaValSerCysVa 404
 :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
 1468 TGTTCCTGCTCTGTGGAGAGGGGATTCAGAGACGGAGCTTTGTGTGTGT 1517
 404 lGluGluAspIleGlnGlyHisValThrSerValGluGluTrpLysCysM 421
 :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
 1518 AGAGAAATCCATGCAATGAGAGAGATTAATCGAGGTGGAAAGAAATGGAAGTGA 1567
 421 eTyrThrProLysMetProIleAlaGlnProCysAsnIlePheAspCys 437
 :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
 1568 TGTACGCACCCAAACCAAGGTTATGCAAACTTGTAAATCTGTGTGATGTC 1617


```

1194 GlyThrLeuSerValLeuLeuHisCysGluAlaIleGlyHisProArgPr 1210
    ||| : : : : : ||| : : : : :
5088 AGGACAGAGGTCATCATATACTGTGTACCTTATCCCCAGTGAGGC 5137
    ||| : : : : : ||| : : : : :
1210 oThrIleSerTrpAlaArgAsnGlyGluValGlnPheSerArgPr 1227
    ||| : : : : : ||| : : : : :
5138 CACATATACATGACCAAGGATGAACCTTGTATACGCCCTCAGTAAAAA 5187
    ||| : : : : : ||| : : : : :
1227 leLeuLeuGlnProAspSerLeuGlnIleLeuAlaProValGluAla 1243
    ||| : : : : : ||| : : : : :
5188 TAATTTGGATGGAACCTGGGAAGATACAGATACAAATCCTCAAGGAAA 5237
    ||| : : : : : ||| : : : : :
1244 AspValGlyPheThrCysAsnAlaThrAsnAlaLeuGlyTyrAspSe 1260
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
5238 GAACAAGGCATATATGATGTCTGTAGTAAATCATCTGTGTTAGATGT 5287
    ||| : : : : : ||| : : : : :
1260 rValSerIleAlaValThrLeuAlaGlyLysProLeuValLysThrSerA 1277
    ||| : : : : : ||| : : : : :
5288 GGAAGTCTCTCTGTGTATGTCAGAGGCACCTGTCACTCTGTGTGTG 5337
    ||| : : : : : ||| : : : : :
1277 rgMetThrValIleAsnThrGluLysProAlaValThrValAspIleGly 1293
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
5338 AAAGAAATATCACCAAAACAGAGCACAAACCATCTGTGTGTGTGGA 5387
    ||| : : : : : ||| : : : : :
1294 SerThrIleLysThrValGlnGlyValAsnValThrIleAsnCysGlnVa 1310
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
5388 GGCATCTGTGAGGAGCCCTTGGAGCAACAGTGACAAATCCGATCTCTGT 5437
    ||| : : : : : ||| : : : : :
1310 lAlaGlyValProGluAlaGluValThrTrpPheArgAsnLysSerLysL 1327
    ||| : : : : : ||| : : : : :
5438 AAAAGGTGTCCTCCAGCCTAATAAATCTGTGTGAGAGGAGATCTC 5487
    ||| : : : : : ||| : : : : :
1327 eu...GlySerProHisHisLeuHisGluLysSerLeuLeuLeuThrAsn 1342
    ||| : : : : : ||| : : : : :
5488 TGAGTGGCAATGTTCTCTCTTCAATGGATCCCTGTTGTTGCAGAAAT 5537
    ||| : : : : : ||| : : : : :
1343 ValSerSerSerAspGlnGlyLeuTyrSerCysArgAlaAlaAsnLeuHi 1359
    ||| : : : : : ||| : : : : :
5538 GTTTCCTTGAAATGAGGAACCTAGCTGTGCATAGCCACCAATGCTCT 5587
    ||| : : : : : ||| : : : : :
1359 sGlyGluLeuThrGluSerThrGlnLeuLeuIleLeuAspProGlnV 1376
    ||| : : : : : ||| : : : : :
5588 TGGAAAGGCAGTGGCAACATCTGTACTCCACATATAAAAGTACATCTCCAG 5637
    ||| : : : : : ||| : : : : :
1376 alProThrGlnLeuGluAspIleArgAla.....LeuLeuAla 1388
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
5638 CAGAGAGTAGAATCGTATTTCTGCAAGGACATATAAAAGTACATCTCCAG 5687
    ||| : : : : : ||| : : : : :
1389 AlaThrGlyProAsnLeuProSerValLeuThrSerProLeuGlyThrGl 1405
    ||| : : : : : ||| : : : : :
5688 GCAACCAACACTAGAACCAACAGC..... 5711
    ||| : : : : : ||| : : : : :
1405 nLeuValLeuAspProGlyAsnSerAlaLeuLeuGlyCysProIleLysG 1422
    ||| : : : : : ||| : : : : :
5712 .....AATGACCCA.....ACAG 5724
    ||| : : : : : ||| : : : : :
1422 lyHisProValProAsnIleThrTrpPheHisGlyGlnProIleVal 1438
    ||| : : : : : ||| : : : : :
5725 GACAACCCCGCCT..... 5738
    ||| : : : : : ||| : : : : :
1439 ThrAlaThrClyLeuThrHisHisIleLeuAlaAlaGlyGlnIleLeuGl 1455
    ||| : : : : : ||| : : : : :
5738 ..... 5738
    ||| : : : : : ||| : : : : :
1455 nValAlaAsnLeuSerGlyGlySerGlnGlyGluPheSerCysLeuAlaG 1472
    ||| : : : : : ||| : : : : :
5738 ..... 5738
    ||| : : : : : ||| : : : : :
1472 lnAsnGluAlaGlyValLeuMetGlnLysAlaSerLeuValIleGlnAsp 1488
    ||| : : : : : ||| : : : : :
5739 .....CAAGAG 5744
    ||| : : : : : ||| : : : : :

```

```

1489 TyrTrpTrpSerValAspArgLeuAlaThrCysSerAlaSerCysGlyAs 1505
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
5745 CTTTTTGGAGCCCTGGTAACCTGGTCACATTTGTTCCACCTGTGTGCA 5794
    ||| : : : : : ||| : : : : :
1505 nArgGlyValGlnGlnProArgLeuArgCysLeuLeu...AsnSerThrG 1521
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
5795 TTTTGGAGCCGCCATTCAGAGACCCCACTGTGTGTGATGGCCAAATGGCAGG 5844
    ||| : : : : : ||| : : : : :
1521 luValAsnProAlaHisCysAlaGlyLysValArgProAlaValGlnPro 1537
    ||| : : : : : ||| : : : : :
5845 AAGTCAGTGAGGCCCTGTGTGATCATCCCTCCAGAACCCACTGTGCTT 5894
    ||| : : : : : ||| : : : : :
1538 lLeAlaCysAsnArgArgAspCysProSerArgTrpMetValThrSerTr 1554
    ||| : : : : : ||| : : : : :
5895 GAGCCCTGTACATCCGGAGCTGCCAGAGGTGGTTCCAAAGTGTGTG 5944
    ||| : : : : : ||| : : : : :
1554 pSerAlaCysThrArgSerCysGlyGlyValGlnThrArgArgValT 1571
    ||| : : : : : ||| : : : : :
5945 GTCACAGTCTCTGTGTCTTGGGTGAAGGATACCACACTCGGCAGGTCA 5994
    ||| : : : : : ||| : : : : :
1571 hrCysGlnLysLeuLysAlaSerGlyIleSerThrProValSerAsnAsp 1587
    ||| : : : : : ||| : : : : :
5995 CGTGCAAGCGGACAAAAGCCAAATGGAACCTGTGCAGGTGTCTCCAAGA 6044
    ||| : : : : : ||| : : : : :
1588 MetCysThrGlnValAlaLysArgProValAspThrGlnAlaCysAsnGl 1604
    ||| : : : : : ||| : : : : :
6045 GCATGTGCCCT...AAAGACGGCCTCTGGGAAGAAACCATGTTTGG 6091
    ||| : : : : : ||| : : : : :
1604 nGlnLeuCysValGluTrpAlaPheSerSerTrpGlyGlnCysAsnGlyP 1621
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
6092 TCATCCATGTGTTCAGTGG.....GAACACAGGAACCGGTGTCTCTGCAC 6135
    ||| : : : : : ||| : : : : :
1621 roCysIleGlyProHisLeuAlaValGlnHisArgGlnValPheCysGln 1637
    ||| : : : : : ||| : : : : :
6136 GTTCATGGGCCGTGTGTGAGGATGCAGACGCTCACACAGCTGTGCAA 6185
    ||| : : : : : ||| : : : : :
1638 ThrArgAspGlyIleThrLeuProSerGluGlnCysSerAlaLeuProAr 1654
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
6186 CACAACAGCTCT.....GACTCCAACCTGTGTGATGACAGAAAGAG 6223
    ||| : : : : : ||| : : : : :
1654 gProValSerThrGlnAsnCysTrpSerGluAlaCysSerValHistTrpA 1671
    ||| : : : : : ||| : : : : :
6224 ACCACCTTAAAGAAGAACTGCATCATCAGGGCCTGTGTGTGTGTGTC 6273
    ||| : : : : : ||| : : : : :
1671 rgValSerLeuTrpThrLeuCysThrAlaThrCysGlyAsnTyrGlyPhe 1687
    ||| : : : : : ||| : : : : :
6274 ACACAGGCCCTTGGAGCCCTGTACAGCAGCCTGTGGCAGG...GGTTC 6320
    ||| : : : : : ||| : : : : :
1688 GlnSerArgArgValGluCysValHisAlaAlaArgThrAsnLysAlaValPr 1704
    ||| : : : : : ||| : : : : :
6321 CAGTCTCGAAAGTCGACTGTATCCACACAGAGGTTGCCAAACCTGTGGC 6370
    ||| : : : : : ||| : : : : :
1704 oGluHisLeuCysSerTrpGlyProArgProAlaAsnTrpGlnArgCysA 1721
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
6371 CAAGAGACACTGTGTACAGAAAAAGAACCAATTTCTCGCGGCACCTGT. 6419
    ||| : : : : : ||| : : : : :
1721 snIleThrProCysGluAsnMetGluCysArgAspThrThrArgTyrCys 1737
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
6420 .CTTGGGCCCTCTCTGTGTAGAGACTGCACAGACACAACTCACTACTGT 6467
    ||| : : : : : ||| : : : : :
1738 GluLysValLysGlnLeuLysLeuCysGlnLeuSerGlnPheLysSerAr 1754
    ||| : : : : : ||| : : : : :
6468 ATGTTTGTAAACATCTTAATTTGTCTCTAGACCGCTACAAACAAAG 6517
    ||| : : : : : ||| : : : : :
1754 gCysCysGlyThrCys 1759
    ||| : : : : : ||| : : : : :
6518 GTGCTGCCAGTCATGT 6533
    ||| : : : : : ||| : : : : :

```

seq_name: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/NA2000.DAT:AAZ98232

seq_documentation_block:

ID AAZ98232 standard; cdna; 3045 bp.

XX


```

267 tAlaGlyProLeuThrAlaAspPheIleValLysIleArgAsnSerGlyS 284
|||||
930 GCGTGGACCACCTACAGCAGATTTCATGTCAGATTCGTAATCGGGCT 979

284 erAlaAspSerThrValGlnPheIlePheTyrGlnProIleIleHisArg 300
|||||
980 CCGCTGACAGTACAGTCCAGTTTCATCTTCTATCAACCCATATCCACCGA 1029

301 TrpArgGluThrAspPheProCysSerAlaThrCysGlyGlyGlyTyr 317
|||||
1030 TGGAGGGAGAGAGGATTTCTTCTCTGCTCAGCAACCTGTGGAGGAGTTA 1079

317 rGlnLeuThrSerAlaGluCysTyrAspLeuArgSerAsnArgValVala 334
|||||
1080 TCAGCTGACATCGGCTGAGTGTACATCTGAGGAGCAACCGTGTGGTTG 1129

334 laAspGlnTyrCysHisTyrTyrProGluAsnIleLysProLysProLys 350
|||||
1130 CTGACCAATACTGTCACTATTACCCAGAGAACATCAAAACCCAAACCAAG 1179

351 LeuGlnGluCysAsnLeuAspProCysProAlaSerAspGlyTyrLysG 367
|||||
1180 CTTTCAAGAGTGCACCTTGATCCCTTGTCCAGCCAGTGACGGATACAAGCA 1229

367 nIleMetProTyrAspLeuTyrHisProLeuProArgTrpGluAlaThrP 384
|||||
1230 GATCATGCTTATGACCTACCATCCCTTCCTCGGTGGAGGCCACCC 1279

384 roTrpThrAlaCysSerSerCysGlyGlyGlyIleGlnSerArgAla 400
|||||
1280 CATGGACCGCGTGTCTCTCTCTGTTGGGGGGGCATCCAGACCGGGCA 1329

401 ValSerCysValGluGluAspIleGlnGlyHisValThrSerValGluG 417
|||||
1330 GTTTCCTGTGGAGGAGGAGCATCCAGGGCATGTCACTTCAGTGAAGA 1379

417 uTrpLysCysMetTyrThrProLysMetProIleAlaGlnProCysAsnI 434
|||||
1380 GTGGAAATGCATGTACACCCCTAGATGCCCATCGCGAGCCCTGCAACA 1429

434 lePheAspCysProLysTrpLeuAlaGlnLutrpSerProCysThrVal 450
|||||
1430 TTTTGTGCTGCCCTAAATGGCTGCACAGGAGTGGTCTCCGTGCACAGTG 1479

451 ThrCysGlyGlnGlyLeuArgTyrArgValValLeuCysIleAspHisAr 467
|||||
1480 ACGTGTGGCCAGGGCCCTCAGATACCGTGGTCTCTGTCATCGACCATCG 1529

467 gGlyMetHisThrGlyGlyCysSerProLysThrLysProHisIleLysG 484
|||||
1530 AGGAATGCACACAGGAGCGTGTAGCCCAAAACAAGACCCCAATAAAG 1579

484 luGluCysIleValProThrProCysTyrLysProLysGluLysLeuPro 500
|||||
1580 AGGAATGCATCGTACCCACCTCCCTGCTATAAACCCAAAGAGAACTTCCA 1629

501 ValGluAlaLysLeuProTrpPheLysGlnAlaGlnGluLutrpSerAlaC 517
|||||
1630 GTCGAGGCCAAGTGGCATGGTTCAAACAAGCTCAAGAGCTAGAAGAAGG 1679

517 yAlaIaValSerGluGluProSerPheIleProGluAlaTrpSerAlaC 534
|||||
1680 AGCTGCGNGTGTACAGAGAGCCCTCGTTCATCCAGAGGCGCTGGTCGGCCT 1729

534 yThrValThrCysGlyValGlyThrGlnValArgIleValArgCysGln 550
|||||
1730 GCACAGTCACCTTNNNTGTGGGACCCAGGTGCGAATAGTCAGTGCCAG 1779

551 ValLeuLeuSerPheSerGlnSerValAlaAspLeuProIleAspGluCy 567
|||||
1780 GTGCTCTGTCTTCTCAAGNNNTGGCTGACCTGCCTATTACGAGAGTG 1829

567 sGluGlyProLysProAlaSerGlnArgAlaCysTyrAlaGlyProCysS 584

```

```

|||||
1830 TGAAGGGCCCAAGCCAGCATCCCGAGGTGCCNNNNNATGCAGGCCCATGCA 1879

584 erGlyGluIleProGluPheAsnProAspGluThrAspGlyLeuPheGly 600
|||||
1880 GGGGGGAAATTCCTGAGTTTCAACCCAGACAGACAGATGGNNNTTGGT 1929

601 GlyLeuGlnAspPheAspGluLeuTyrAspTrpGluTyrGluGlyPheTh 617
|||||
1930 GGCTGCAGGATTCGACGAGCTGTATGACTGGGAGTATGAGGGGTTCCAC 1979

617 rLysCysSerGluSerCysGlyGlyValGlnGluAlaValValSerC 634
|||||
1980 CAANNCTCCGAGTCTGTGGAGGAGGTGCCAGAGGCTGTGGTGAGCT 2029

634 yLeuAsnLysGlnThrArgGluProAlaGluGluAsnLeuCysValTh 650
|||||
2030 GCTTGAACAACAAANNCTCGGAGCCCTGTGAGGAGCAACCTGTGCGTGAC 2079

650 rSerArgArgProGlnLeuLysSerCysAsnLeuAspProCysP 667
|||||
2080 CACCGCGGCCCCACACAGCTCTGAGTCCCGCAATTTGGATGCCCTGCC 2128

667 roAla..ArgTrpGluIleGlyLys.....TipSerP 677
|||||
2129 CAGCAAGTCTCTCATCTAGGAAGAGCAGTATCGACTCAGCATGGAACG 2178

677 roCysSerLeuThrCysGlyValGlyLeuGlnThrArgAspValPheCys 693
|||||
2179 CTTGCAACGCTCTTGTGTTAGGCAACCAAGAGCGCTGCTTCATCCTGCC 2228

694 SerHisLeuLeuSer..... 698
|||||
2229 TGTCAACCACTAGCTCTGTGGCTAGGGGAGGTGTCGCCCTTATGTT 2278

699 ....ArgGluMetAsnGluThrValIleLeuAlaAspGluLeuCysArgG 714
:::
2279 TCCACATCTGCAAAAGTGAACCTGGTTGTA..... 2306

714 lnProLysProSerThrValGlnAlaCysAsnArgPheAsnCysProPro 730
|||||
2307 .....CCTGATGATCTGATGCCATGCCATCTGCTCACATGTCCTCATG 2348

731 AlaTrpTyrProAlaGlnTrpGlnProCysSerArgThrCysGlyGlyCl 747
|||||
2349 ATTCTTAT.....TTGTAGGCAGAAAG 2371

747 yVal...GlnLysArgGluValLeuCysLysGlnArgMetAlaAspGlyS 763
:::
2372 CATTAACACAGCTACTCCGCTGCTGTGCTAATCATTCCTGTAAT.... 2417

763 erPheLeuGluLeuProGluThrPheCysSerAlaSerLysProAlaCys 779
|||||
2418 .....TTCTGTCTCTGCTTATTTGCCATTATT 2444

780 GlnGlnAlaCysLysLysAspCysProSerGluTrpLeuLeuSerAs 796
:::
2445 GAAACACATGCAAAAGGCTCTTTTACCACATTCCTGTTGTAACA... 2492

796 pTrpThrGluCysSerThrSerCys...GlyGluGlyThrGlnThrArgS 812
:::
2493 .....ACACCCAAATGCTGAGGAGTCCGCGAGGAGTCAGTGC 2529

812 erAlaIleCysArgLysMetLeuLysThrGlyLeuSerThrValValAsn 828
:::
2530 CTGGGAGTCTGCTTAAAACTGCTG..... 2552

829 SerThrLeuLeuProLeuProPhe 837
|||||
2553 ...GGACTCGTGGTCCCTAAACCCCTTC 2576

```

seq_name: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA2001A.DAT:ABA06676

see documentation block:

PR 01-DEC-2000; 2000US-250391P.
 PR 05-DEC-2000; 2000US-251030P.
 PR 05-DEC-2000; 2000US-251988P.
 PR 05-DEC-2000; 2000US-256719P.
 PR 06-DEC-2000; 2000US-251479P.
 PR 08-DEC-2000; 2000US-251856P.
 PR 08-DEC-2000; 2000US-251868P.
 PR 08-DEC-2000; 2000US-251869P.
 PR 08-DEC-2000; 2000US-251989P.
 PR 08-DEC-2000; 2000US-251990P.
 PR 11-DEC-2000; 2000US-254097P.
 PR 05-JAN-2001; 2001US-259678P.
 XX
 PA (HUMA-) HUMAN GENOME SCI INC.
 XX
 XX Rosen CA, Barash SC, Ruben SM;
 XX
 DR WPI; 2001-476161/51.
 DR P-PSDB; ABB10454.
 XX
 PT Isolated nucleic acid molecule encoding an inflammation-associated
 PT polypeptide is used in preventing, treating or ameliorating a medical
 PT condition -
 XX
 PS Claim 1; SEQ ID NO: 342; 859pp + Sequence Listing; English.
 XX
 CC The present invention provides human cDNAs, proteins and related genomic
 CC DNAs. These can be used in the treatment of neural, immune system,
 CC muscular, reproductive, gastrointestinal, pulmonary, cardiovascular,
 CC renal and proliferative disorders and inflammation. The present sequence
 CC is a cDNA of the invention.
 XX
 SQ Sequence 2031 BP; 449 A; 610 C; 564 G; 408 T; 0 other;

alignment_scores:

Quality: 3423.00 Length: 637
 Ratio: 5.374 Gaps: 0
 Percent Similarity: 100.000 Percent Identity: 99.843

alignment_block:

US-10-044-807-2 x ABA06676 ..
 Align seg 1/1 to: ABA06676 from: 1 to: 2031
 1126 ArgThrSerProValThrLeuSerProHisLysHisValSerGlyPhe 1142
 27 AGGACTTCCCGAGTGACTCTCTCGCTCATAAACAGCTGTCTGCTCAG 76
 1142 rSerSerLeuArgThrSerSerThrGlyAspAlaGlyGlySerArg 1159
 77 CAGCTCCCTGGGACCTCTCTCCACCGGGGACGCGGGGAGGCTCTCGAA 126
 1159 rgProHisArgLysProThrIleLeuArgLysIleSerAlaAlaGlnGln 1175
 127 GGCCACACCGCAGGCCACCATCTCTCGCAAGATCTCAGGGCCACGAG 176
 1176 LeuSerAlaSerGluValValThrHisLeuGlyGlnThrValAlaLeuAl 1192
 177 CTCTCAGCCTCGAGGTGTCTCTCCACCGGGGACGCGGGGAGGCTCTG 226
 1192 aSerGlyThrLeuSerValLeuHisCysGluAlaIleGlyHisProAl 1209
 227 CAGCGGACACTCAGTGTCTCTCTGACATGTGAGGCCATCGGCCACCAA 276
 1209 rgProThrIleSerTrpAlaArgAsnGlyGluGluValGlnPheSerAsp 1225
 277 GGCTACCATCAGTGGGCCAGGATGGAGAGAGATTCAGTTCAGTGAC 326
 1226 ArgIleLeuLeuGlnProHisPheSerLeuGlnIleLeuAlaProValG 1242
 327 AGGATTCTTACAGCCAGATGATTCCTTACAGATCTTGGCACCAGTGCA 376

1242 uAlaaspValGlyPheTyrThrCysAsnAlaThrAsnAlaLeuGlyTyrA 1259
 377 AGCAGATGTGGGTTTCTACACTTGCATGCCACCAATGCTTGGGATACG 426
 1259 spSerValSerIleAlaValThrLeuAlaGlyLysProLeuVallyThr 1275
 427 ACTCTGTCTCCATTGCCGTACATTAGCAGGAAGCCACTAGTGAACAG 476
 1276 SerArgMetThrValIleasnThrGluLysProAlaValThrValAsp 1292
 477 TCACCAATGCAGTGTATCAACAGGAGAGCTGTCAGTCACAGTCGATAT 526
 1292 eGlySerThrIleLysThrValGlnGlyValAsnValThrIleAsnCysG 1309
 527 AGGAAGCACCATCAAAACAGTGCAGGGAGTGAATGTGACAAATCAACTGCC 576
 1309 InValAlaGlyValProGluAlaGluValThrTrpPheArgAsnLysSer 1325
 577 AGGTTGCAGAGTGCCTGAAGCTGAAGTCACCTTGGTTCAGGATATAAGC 626
 1326 LysLeuGlySerProHisHisLeuHisGluGlySerLeuLeuThrAs 1342
 627 AAACCTGGGCTCCCGCACCATCTGCACGAAGGCTCCTTGTCTCACA 676
 1342 nValSerSerSerAspGlnGlyLeuTyrSerCysArgAlaAlaAsnLeuH 1359
 677 CGTGTCTCTCGATCAGGAGGCTGACTCTGAGGGGCGGCAATCTTC 726
 1359 isGlyGluLeuThrGluSerThrGlnLeuLeuIleLeuAspProGln 1375
 727 ATGGAGAGCTGACTGAGAGCACCCAGCTGCTGATCCTAGATCCCCCA 776
 1376 ValProThrGlnLeuGluAlaAspIleArgAlaLeuAlaAlaThrGlyPr 1392
 777 GTCCACACAGTTGGAGACATCAGGGCTTGTCTGCTGCTGCTGAC 826
 1392 oAsnLeuProSerValLeuThrSerProLeuGlyThrGlnLeuValLeuA 1409
 827 GAACCTTCCCTCAGTGTCTGCTCTCTGCGGACACAGCTGGTCTG 876
 1409 spProGlyAsnSerAlaLeuLeuGlyCysProIleLysGlyHisProVal 1425
 877 ATCTGGGAATTCGTCTCTTGGCTGCCCATCAAAAGTCACTCCCTGTC 926
 1426 ProAsnIleThrTrpPheHisGlyGlnProIleValThrAlaThrGl 1442
 927 CCAATATCACCTGGTTTCATGGTGTGTCAGCCAAATTCCTCAGTCCACAG 976
 1442 YLeuThrHisHisIleLeuAlaAlaGlyGlnIleLeuGlnValAlaAsnL 1459
 977 ACTGACGCATCACATCTTGGCAGCTGGACAGATCCTTCAAGTTGCAAA 1026
 1459 euSerGlyGlySerGlnGlyGluPheSerCysLeuAlaGlnAsnGluAla 1475
 1027 TTAGCGGTGGGTCTCAAGGGGAATTCAGCTGCTTGTCTCAGATGAGGCA 1076
 1476 GlyValLeuMetGlnLysAlaSerLeuValIleGlnAspTyrTrpPse 1492
 1077 GGGGTGCTCATGCAAGAGGCATCTTTAGTGTATCCAAAGATTACTGGTGC 1126
 1492 rValAspArgLeuAlaThrCysSerAlaSerCysGlyAsnArgGlyValG 1509
 1127 TGTGGACAGACTGGCAACCTGCTCAGCTTCTGTGTGTACCGGGGGTTC 1176
 1509 InGlnProArgLeuArgCysLeuLeuAsnSerThrGluValAlaAsnProAla 1525
 1177 AGCAGCGCGCTTGGGTGCTGCTGAACAGCAGCAGGAGGTCAACCTGCC 1226
 1526 HisCysAlaGlyLysValArgProAlaValGlnProIleAlaCysAsnAr 1542
 1227 CACTGGCAGGGAAGGTTCGCTTGTGGTGCAGCCCATCGCGTCAACCG 1276
 1542 gArgAspCysProSerArgTrpMetValThrSerTrpSerAlaCysThra 1559

PR 13-MAY-1998; 98US-0085338.
 PR 13-MAY-1998; 98US-0085339.
 PR 15-MAY-1998; 98US-0085573.
 PR 15-MAY-1998; 98US-0085579.
 PR 15-MAY-1998; 98US-0085580.
 PR 15-MAY-1998; 98US-0085582.
 PR 15-MAY-1998; 98US-0085689.
 PR 15-MAY-1998; 98US-0085697.
 PR 15-MAY-1998; 98US-0085700.
 PR 15-MAY-1998; 98US-0085704.
 PR 18-MAY-1998; 98US-0086023.
 PR 22-MAY-1998; 98US-0086392.
 PR 22-MAY-1998; 98US-0086414.
 PR 22-MAY-1998; 98US-0086430.
 PR 22-MAY-1998; 98US-0086486.
 PR 28-MAY-1998; 98US-0087098.
 PR 28-MAY-1998; 98US-0087106.
 PR 28-MAY-1998; 98US-0087208.
 PR 30-JUL-1998; 98US-0094651.
 PR 11-SEP-1998; 98US-0100038.
 XX
 PA (GETH) GENENTECH INC.
 XX
 PI Wood WI, Goddard A, Gurney A, Yuan J, Baker KP, Chen J;
 XX
 DR WPI: 1999-551358/46.
 DR P-PSDB; AA41729.
 XX
 PT New secreted and transmembrane polypeptides and their polynucleotides,
 PT useful for treating blood coagulation disorders, cancers and cellular
 PT adhesion disorders -
 XX
 PS Claim 2; Fig 119; 530pp; English.
 XX
 CC The present invention describes secreted and transmembrane polypeptides
 CC and their polynucleotides. The nucleotide sequences are useful as
 CC sources of probes, primers, for chromosome mapping, and for generation
 CC of antisense sequences. They can also be used to create transgenic
 CC animals. The proteins can be used to treat a variety of diseases and
 CC disorders, depending on their function. Diseases that may be treated
 CC include blood coagulation disorders, cancers and cellular adhesion
 CC disorders. They may also be used to raise antibodies. AA233891 to
 CC AA234338, and AA41685 to AA41774 represent polynucleotide and
 CC polypeptide sequence given in the exemplification of the present
 CC invention.
 XX
 SQ Sequence 1869 BP; 490 A; 469 C; 480 G; 430 T; 0 other;

alignment_scores:
 Quality: 2919.00 Length: 530
 Ratio: 5.528 Gaps: 0
 Percent Similarity: 99.623 Percent Identity: 99.057

alignment_block:
 US-10-044-807-2 x AA234146 ..

Align seg 1/1 to: AA234146 from: 1 to: 1869

1 MetGluCysCysArgAlaThrProGlyThrLeuLeuPheLeuAl 17
 |||||
 133 ATGGAAATGTCGCGTGGGCAACTCTGSCACACACTCTCTTCTTGGC 182

17 aPheLeuLeuSerSerArgThrAlaArgSerGluGluAspArgAspG 34
 |||||
 183 TTTCCTGCTCTCTGAGTTCAGACCGCAGCTCCGAGGAGGACCGGAGC 232

34 lYLeuTrpAspAlaTrpGlyProTrpSerGluCysSerArgThrCysGly 50
 |||||
 233 GCCTATGGATGCTCTGGGCCCATCGAGTGAATGCTCAGCACCTGCGGG 282

51 GlyGlyAlaSerTyrSerLeuArgArgCysLeuSerSerLysSerCysG 67
 |||||

283 GGAGGGGCTCTACTCTCTGAGCGCTGCTGAGCAGCAAGAGCTGTGA 332
 67 uGlyArgAsnIleArgTyrArgThrCysSerAsnValAspCysProProG 84
 |||||
 333 AGGAAGAAATATCCGATACAGAACATGCACTAATGTGGACTGCCACCA 382
 84 luAlaGlyAspPheArgAlaGlnGlnCysSerAlaHisAsnAspValLys 100
 |||||
 383 AAGCAGGTGATTTCCGAGCTCAGCAATGCTCAGCTCATAATGATGTCA 432
 101 HisHisGlyGlnPheTyrGluTrpLeuProValSerAsnAspProAspAs 117
 |||||
 433 CACCATGGCCAGTTTATGAATGGCTTCTGTGTCTAATGACCCCTGACA 482
 117 nProCysSerLeuLysCysGlnAlaLysGlyThrThrLeuValValGluL 134
 |||||
 483 CCCATGTTCACTCAAGTCCCAAGCCAAAGCAACACCTGGTGTGTGAAC 532
 134 euAlaProLysValLeuAspGlyThrArgCysTyrThrGluSerLeuAsp 150
 |||||
 533 TAGCACCTAAGTCTTAGATGGTACGCTGCTATACAGAACTCTTTGGAT 582
 151 MetCysIleSerGlyLeuCysGlnIleValGlyCysAspHisGlnLeuGl 167
 |||||
 583 ATGTGCATCAGTGTGTTATGCCAAATTTGTCCTGCGATCACCAGCTGG 632
 167 ySerThrValLysGluAspAsnCysGlyValCysAsnGlyAspGlySert 184
 |||||
 633 AAGCACCGTCAAGGAAGATAACTGTGGGCTGCAACGGAGATGGGTCCA 682
 184 hrCysArgLeuValArgGlyGlnTrpLysSerGlnLeuSerAlaThrLys 200
 |||||
 683 CCTCCGCTGGTCCGAGGCGAGTATAATCCAGCTCTCCGCAACCAAA 732
 201 SerAspAspThrValValAlaIleProTyrGlySerArgHisIleArgLe 217
 |||||
 733 TCGGATGATACTGTGGTTGCATCTCCCTATGGAAGTAGACATATTCGCT 782
 217 uValLeuLysGlyProAspHisLeuTyrLeuGluThrLysThrLeuGlnG 234
 |||||
 783 TGTCTTAAAAAGGCTCTGATCACTATATCTGGAACCAACCAACCTCCAG 832
 234 lYThrLysGlyGluAsnSerLeuSerSerThrGlyThrPheLeuValAsp 250
 |||||
 833 GGACTAAAGTGAACAGTCTCAGCTCCAGGAACTTCTCTGTGGAC 882
 251 AsnSerSerValAspPheGlnLysPheProAspLysGluIleLeuArgMe 267
 |||||
 883 AATTCTAGTGTGGACTTCCAGAAATTTCCAGACAAAGAGATAGAGAAT 932
 267 tAlaGlyProLeuThrAlaAspPheIleValLysIleArgAsnSerGlyS 284
 |||||
 933 GGTGGACCACTCAGCAGAGATTTCATGTCAAGATTTCGTAACCTCGGCT 982
 284 eAlaAspSerThrValGlnPheIlePheTyrGlnProIleIleHisArg 300
 |||||
 983 CCGCTCAGAGTACAGTCCAGTCACTCTTCTATCAACCCATCATCCACCG 1032
 301 TrpArgGluThrAspPhePheProCysSerAlaThrCysGlyGlyGlyTy 317
 |||||
 1033 TGGAGGGAGACGATTTCTTCTCTGCTCAGCAACCTGTGGAGGAGTTA 1082
 317 rGlnLeuThrSerAlaGluCysTyrAspLeuArgSerAsnArgValValA 334
 |||||
 1083 TCAGCTGACATCGGCTGAGTGTACGATCTGAGGAGCAACCTGTGGTGTG 1132
 334 laAspGlnTrpCysHisTyrTyrProGluAsnIleLysProLysProLys 350
 |||||
 1133 CTGACCAATACTGTCACTATTACCCAGAGAAATCAAAACCCCAACCAAG 1182
 351 LeuGlnGluCysAsnLeuAspProCysProAlaSerAspGlyTyrLysG 367
 |||||
 1183 CTTTCAGAGTGCACACTTGGATCTTCTGTCAGCCAGTACGGATACAGCA 1232

```

367 nileMetProTyrAspIeuTyrHisProLeuProArgTrrpGluAlaThrP 384
1233 GATCATGCTTATGACCTTACCATCCCTCCCTCGGTGGAGGCCACCC 1282
384 roTrrpThraLacysSerSerCysGlyGlyGlyIleGlnSerArgAla 400
1283 CATGGACCGGCTGCTCTCTGCTGGGGGGGCATCCAGAGCGGGCA 1332
401 ValSerCysValGluGluAspIleGlnGlyHisValThrSerValGluG 417
1333 GTTTCCTGCTGGAGGAGGACATCCAGGGGCATGCTCACTTCAGTGAAGA 1382
417 uTrrpLysCysMetTrrpThrProLysMetProIleAlaGlnProCysAsnI 434
1383 GTGGAAATGCATGTACACCCCTAAGATGCCCTATCGCGAGCCCTGCNACA 1432
434 lePheAspCysProLysTrrpLeuAlaGlnLutrrpSerProCysThrVal 450
1433 TTTTTCACCTGCTAAATGGCTGGCAGCAGGTGGTCTCCGTGCACAGTG 1482
451 ThrCysGlyGlnGlyLeuArgTrrpArgValValLeuCysIleAspHisAr 467
1483 ACATGTGGCGAGGGCTCAGATACCTGTGGTCTCTGTCATCGACCATCG 1532
467 gGlyMetHisThrGlyGlyCysSerProLysThrLysProHisIleLysG 484
1533 AGGAATGCACACAGGAGGCTGTAGCCCAAAACAAAGCCCAATAAAG 1582
484 luGluCysIleValProThrProCysTrrpLysProLysGluLysLeuPro 500
1583 AGCAATGCATCGTACCCACCTCCCTGCTATAAACCAAGAGAACTTCCA 1632
501 ValGluAlaLysLeuProTrrpPheLysGlnAlaGlnLutrrpGluGlu 517
1633 GTCGAGGCCAAGTGGCATGGTTCACAAACAGCTCAAGAGCTAGAAGAGG 1682
517 YAlaAlaValSerGluGluProSerPheIleProGluAla 530
1683 AGCTGTGTGTGACAGAGGCCCTCGTAAAGTGTAAAGCA 1722

seq_name: /SIN1/gcgdata/hold-geneseq/geneseqn-emb1/NA2000.DAT.AAC78532
seq_documentation_block:
ID AAC78532 standard; cDNA; 1869 BP.
XX AAC78532;
XX
XX 08-FEB-2001 (first entry)
XX
XX Human PRO1071 (UNQ528) nucleotide sequence SEQ ID NO:300.
XX
XX Human; secreted protein; transmembrane protein; PRO; EST; cytotstatic;
KW expressed sequence tag; detection; cancer; ss.
XX
XX Homo sapiens.
XX
XX WO200053756-A2.
XX
XX 14-SEP-2000.
XX
XX 18-FEB-2000; 2000WO-US04341.
XX
XX 08-MAR-1999; 99WO-US05028.
XX
XX 12-MAR-1999; 99US-0123957.
XX
XX 29-MAR-1999; 99US-0126773.
XX
XX 21-APR-1999; 99US-0130232.
XX
XX 28-APR-1999; 99US-0131445.
XX
XX 14-MAY-1999; 99US-0134287.
XX
XX 23-JUN-1999; 99US-0141037.
XX
XX 26-JUL-1999; 99US-0145698.
XX
XX 29-OCT-1999; 99US-0162506.
XX
XX 30-NOV-1999; 99WO-US28313.

102-DEC-1999; 99WO-US28551.
102-DEC-1999; 99WO-US28555.
16-DEC-1999; 99WO-US30095.
30-DEC-1999; 99WO-US31243.
30-DEC-1999; 99WO-US31274.
05-JAN-2000; 2000WO-US00219.
06-JAN-2000; 2000WO-US00277.
06-JAN-2000; 2000WO-US00376.
XX
XX (GETH ) GENENTECH INC.
XX
XX Ashkenazi AJ, Baker KP, Botstein D, Desnoyers L, Eaton DL;
XX Ferrara N, Filvaroff E, Fong S, Gao W, Gerber H, Gerritsen ME;
XX Goddard A, Godowski PJ, Grimaldi CJ, Gurney AL, Hillan KJ;
XX Kljavin IJ, Kuo SS, Napier MA, Pan J, Paoni NF, Roy MA;
XX Shelton DL, Stewart TA, Tumas D, Williams PM, Wood WI;
XX
XX WPI; 2000-611443/58.
XX P-PSDB; AAB44285.
XX
XX Novel PRO polypeptides and polynucleotides used in detection methods,
XX to target bioactive molecules to specific cells, and to modulate
XX cellular activities -
XX
XX Claim 2; Fig 119; 636pp; English.
XX
XX AAC78458 to AAC78599 represent polynucleotide and EST (expressed
XX sequence tag) sequences which encode secreted or transmembrane PRO
XX polypeptides. The PRO polynucleotides and polypeptides have cytostatic
XX activity. The polynucleotides and polypeptides can be used for detecting
XX the presence of PRO polypeptides in samples, for linking bioactive
XX molecules to cells and for modulating biological activities of cells,
XX using the polypeptides for specific targeting. The polypeptide targeting
XX can be used to kill the target cells, e.g. for the treatment of cancers.
XX The polypeptide pairs provide specific targeting of bioactive molecules
XX to cells. AAC78600 to AAC78987 represent PCR primers and probes used in
XX the isolation of the PRO polynucleotide sequences.
XX
XX Sequence 1869 BP; 490 A; 469 C; 480 G; 430 T; 0 other;

alignment_scores:
Quality: 2919.00 Length: 530
Ratio: 5.528 Gaps: 0
Percent Similarity: 99.623 Percent Identity: 99.057
alignment_block:
US-10-044-807-2 x AAC78532 ..
Align seg 1/1 to: AAC78532 from: 1 to: 1869
1 MetGluCysCysArgArgAlaThrProGlyThrLeuLeuPheLeuAl 17
133 ATGGAATGCTGCGTGGGCACTCTGGCACACTGCTCTCTTTCTGGC 182
17 aphetLeuLeuSerSerArgTrrpAlaArgSerGluGluAspArgAspG 34
183 TTTCTGCTCTGAGTTCAGACCGCAGCTCCGAGGAGGACCGGAGC 232
34 lYLeuTrrpAspAlaTrrpGlyProTrrpSerGluCysSerArgThrCysGly 50
233 GCCATGGATGCTCGGGGCCCATGGAGTGAATGCTCAGCACCCTCGCGG 282
51 GlyGlyAlaSerTyrSerLeuArgArgCysLeuSerSerLysSerCysG 67
283 GGAGGGGCTCTCTACTCTCTGAGGCGCTGCTGAGCAGCAGAGCTGTGA 332
67 uGlyArgAsnIleArgTrrpArgThrCysSerAsnValAspCysProProG 84
333 AGGAGAGAAATATCCGATACAGAACATGCAGTAAATGTGGACTGCCACCA 382
84 luAlaGlyAspPheArgAlaGlnGlnCysSerAlaHisAsnAspValLys 100
```

```

383 AAGCAGGTGATTTCGAGCTCAGCAATGCTCAGCTCATATAATGATGTCAG 432
101 HisHisGlyGlnPheTyrGluThrLeuProValSerAsnAspProAspAs 117
433 CACCATGGCCAGTATTTATGATGCTTCTCTGCTGCTAATGACCTGACAA 482
117 nProCysSerLeuLysCysGlnAlaLysGlyThrThrLeuValValGluL 134
483 CCCATGTTCACTCAAGTGCCCAAGGCAAGGCAACACCTGGTTGTTGAAC 532
134 euAlaProLysValLeuAspGlyThrArgCysTyrThrGluSerLeuAsp 150
533 TAGCACCTTAAGGCTCTAGATGGTACGGTGTCTATACAGAACTTTGGAT 582
151 MetCysIleSerGlyLeuCysGlnIleValGlyCysAspHisGlnLeuG 167
583 ATGTGTCATCAGTGGTATATGCCAAATGTTGGTGGCATCACCAGCTGGG 632
167 ySerThrValLysGluAspAsnCysGlyValCysAsnGlyAspGlySerT 184
633 AAGCACCTCAAGGAAGATACCTGTGGGCTGCAACGGAGATGGGTCCA 682
184 hrCysArgLeuValArgGlyGlnTyrLysSerGlnLeuSerAlaThrLys 200
683 CCTGCCGGCTGGTCCGAGGCGAGTATAATCCAGCTCTCCGCAACCAAA 732
201 SerAspAspThrValValAlaIleProTyrGlySerArgHisIleArgLe 217
733 TCGATGATGACTGTGGTGTGACATCTCCCTATGGAAGTAGACATATTCCT 782
217 uValLeuLysGlyProAspHisLeuTyrLeuGluThrLysThrLeuGlnG 234
783 TGTCTTAAAGGTCCCTGATCATTATATCTGGAACCAAAACCCCTCCAGG 832
234 lyThrLysGlyGluAsnSerLeuSerThrGlyThrPheLeuValAsp 250
833 GGACTAAAGGTGAAACAGCTCTCAGCTCCACAGGAACCTTCCTTGTGGAC 882
251 AsnSerSerValAspPheGlnLysPheProAspLysGluIleLeuArgMe 267
883 AATCTAGTGTGGACTTCCAGAAATTTCCAGACAAAGAGATCTGAGAA 932
267 tAlaGlyProLeuThrAlaAspPheIleValLysIleArgAsnSerGlyS 284
933 GGCTGGACCACTCACAGCAGATTCTAATGTCAAGATTCTGTAACCTCGGCT 982
284 exAlaAspSerThrValGlnPheIlePheTyrGlnProIleIleHisArg 300
983 CCGCTGACAGTACAGTCCAGTTCATCTTCTATCAACCCCATCATCCACGA 1032
301 TrpArgGluThrAspPhePheProCysSerAlaThrCysGlyGlyGlyTy 317
1033 TGGAGGGAGACGGATTCTTCTTCTGCTCAGAACCTGTGGAGGAGGTAA 1082
317 rGlnLeuThrSerAlaGluCysTyrAspLeuArgSerAsnArgValVala 334
1083 TCAGCTGACATCGGCTGAGTGTACGATCTGAGGACCAACCGTGTGGTTG 1132
334 laAspGlnTyrCysHisTyrTyrProGluAsnIleLysProLysProLys 350
1133 CTGACCAATAGTGTCACTATTACCCAGAGAATCAAAACCCCAACCCCAAG 1182
351 LeuGlnGluCysAsnLeuAspProCysProAlaSerAspGlyTyrLysG 367
1183 CTTAGGAGTGGCACTTGGATCTTGTCCAGCCAGTGCAGGATACAAGCA 1232
367 nIleMetProTyrAspLeuTyrHisProLeuProArgTrpGluAlaThrP 384
1233 GATCATGCTTATGACCTCTACCATCCCTTCTCTCGTGGGAGGCCACCC 1282
384 roTrpThrAlaCysSerSerSerCysGlyGlyGlyIleGlnSerArgAla 400
1283 CATGACCGCGTCTCTCTCTCTGTTGGGGGGGATCCAGAGCCGGCA 1332

```

```

401 ValSerCysValGluGluAspIleGlnGlyHisValThrSerValGluG 417
1333 GTTTCCTGTGTGGAGGAGGACATCCAGGGGCATGCTACTTCAGTGGAGA 1382
417 uTrpLysCysMetTyrThrProLysMetProIleAlaGlnProCysAsnI 434
1383 GTGAAATGCATGTACACCCCTAGATGCCATCGCGAGCCCTGCAACA 1432
434 lePheAspCysProLysTrpLeuAlaGlnGluTrpSerProCysThrVal 450
1433 TTTTGTACTGCCTAAATGGCTGGCACAGAGTGTCTCCGTGCACAGTG 1482
451 ThrCysGlyGlnGlyLeuArgTyrArgValValLeuCysIleAspHisAr 467
1483 ACATGTGGCCAGGGCTTCAGATACCGTGTGTGCTCTGTGCATCGACCA 1532
467 gGlyMetHisThrGlyGlyCysSerProLysThrLysProHisIleLysG 484
1533 AGGAATGCACACAGSAGGCTGTAGCCCAAAACAAAGCCCAATAAAG 1582
484 luGluCysIleValProThrProCysTyrLysProLysGluLysLeuPro 500
1583 AGGAATGCATCGTACCCACTCCCTGCTATAAAACCCAAAGAGAACTTCCA 1632
501 ValGluAlaLysLeuProThrPheLysGlnAlaGlnGluLeuGluG 517
1633 GTCAGGCCCAAGTTGCCATGTTCAACAAGCTCAAGAGCTAGAAGAAG 1682
517 yAlaAlaValSerGluProSerPheIleProGluAla 530
1683 AGCTGCTGTGTGAGAGGAGCCCTCGTAAGTTGTAAAGCA 1722

seq name: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/NA2001A.DAT:AAS46142
seq_documentation_block:
ID AAS46142 standard; cDNA; 1869 bp.
XX
AC AAS46142;
XX
DT 18-DEC-2001 (first entry)
XX
DE Human DNA encoding PRO polypeptide sequence #218.
XX
KW PRO polypeptide; mammal; tumour; cancer; human; cattle; horse; sheep; ss;
KW dog; cat; pig; goat; rabbit; tumour necrosis factor alpha; TNF-alpha;
KW blood; chondrocyte cell; cell proliferation; cell differentiation; colon;
KW adrenal; lung; breast; prostate; rectum; cervix; liver; genetic disorder;
KW PCR primer.
XX
OS Homo sapiens.
XX
PN WO200168848-A2.
XX
PD 20-SEP-2001.
XX
PF 28-FEB-2001; 2001WO-US06520.
XX
PR 01-MAR-2000; 2000WO-US05601.
PR 02-MAR-2000; 2000WO-US05841.
PR 03-MAR-2000; 2000US-187202P.
PR 06-MAR-2000; 2000US-186968P.
PR 14-MAR-2000; 2000US-189320P.
PR 14-MAR-2000; 2000US-189320P.
PR 15-MAR-2000; 2000WO-US06884.
PR 21-MAR-2000; 2000US-190828P.
PR 21-MAR-2000; 2000US-191007P.
PR 21-MAR-2000; 2000US-191048P.
PR 28-MAR-2000; 2000US-191314P.
PR 28-MAR-2000; 2000US-192655P.
PR 29-MAR-2000; 2000US-193032P.
PR 30-MAR-2000; 2000US-193053P.
PR 30-MAR-2000; 2000WO-US08439.

```

PR 04-APR-2000; 2000US-194449P.
 PR 04-APR-2000; 2000US-194647P.
 PR 11-APR-2000; 2000US-195975P.
 PR 11-APR-2000; 2000US-196000P.
 PR 11-APR-2000; 2000US-196187P.
 PR 11-APR-2000; 2000US-196690P.
 PR 11-APR-2000; 2000US-196820P.
 PR 18-APR-2000; 2000US-198121P.
 PR 18-APR-2000; 2000US-198585P.
 PR 25-APR-2000; 2000US-199397P.
 PR 25-APR-2000; 2000US-199550P.
 PR 25-APR-2000; 2000US-199654P.
 PR 03-MAY-2000; 2000US-201516P.
 PR 17-MAY-2000; 2000WO-US13705.
 PR 22-MAY-2000; 2000WO-US14042.
 PR 30-MAY-2000; 2000WO-US14941.
 PR 02-JUN-2000; 2000WO-US15284.
 PR 05-JUN-2000; 2000US-209832P.
 PR 28-JUL-2000; 2000WO-US20710.
 PR 22-AUG-2000; 2000US-0644848.
 PR 24-AUG-2000; 2000WO-US23328.
 PR 08-NOV-2000; 2000WO-US30952.
 PR 01-DEC-2000; 2000WO-US32678.
 PR 20-DEC-2000; 2000WO-US34956.
 XX
 PA (GETH) GENENTECH INC.

XX Baker KP, Chen J, Desnoyers L, Goddard A, Godowski PJ, Gurney AL;
 PI Pan J, Smith V, Watanabe CK, Wood WI, Zhang Z;
 XX
 XX WPI; 2001-602746/68.
 DR P-PSDB; AAU29241.

XX Novel nucleic acids encoding PRO polypeptides, used to diagnose the
 PT presence of tumours, such as prostate and breast tumours, in mammals and
 PT to screen for modulators of the compounds -

XX Claim 2; Fig 435; 774pp; English.

XX Sequences AAS45925-AAS46231 represent DNA molecules encoding and PCR
 CC primers for PRO polypeptides of the invention. The sequences of the
 CC invention can be used to detect the presence of a tumour in a mammal by
 CC comparing the level of expression of a PRO polypeptide in a test sample
 CC of cells from the animal and a control sample of normal cells, whereby a
 CC higher level of expression in the test sample indicates the presence of a
 CC tumour in the mammal. Mammals include dogs, cats, cattle, horses, sheep,
 CC pigs, goats and rabbits but are preferably human. The polypeptides can be
 CC used to stimulate tumour necrosis factor (TNF) alpha release from human
 CC blood, when contacted with it. A specific polypeptide can be used to
 CC stimulate the proliferation or differentiation of chondrocyte cells. The
 CC PRO proteins can be used to determine the presence of tumours and also
 CC susceptibility to tumour development, particularly adrenal, lung, colon,
 CC breast, prostate, rectal, cervical, or liver tumours, in mammalian
 CC subjects. The oligonucleotide probes specific for the PRO nucleic acids
 CC can be used for genetic analysis of individuals with genetic disorders.

XX Sequence 1869 BP; 490 A; 469 C; 480 G; 430 T; 0 other;

alignment_scores:
 Quality: 2919.00 Length: 530
 Ratio: 5.528 Gaps: 0
 Percent Similarity: 99.623 Percent Identity: 99.057

alignment_block:

US-10-044-807-2 x AAS46142 ..

Align seg 1/1 to: AAS46142 from: 1 to: 1869

1 MetGluCysCysArgAlaThrProGlyThrLeuLeuLeuLeuAl 17
 |||||
 133 ATGGATGCTGCCGTCGGGCAACTCCGGCACACTGCTCCTTCTTCTGGC 182

17 aPheLeuLeuLeuSerSerArgThrAlaArgSerGluGluAspArgAspG 34
 |||||
 183 TTTCTCTGCTCTCAGTCCAGGACCGCAGCTCCGAGGAGGACCGGGGACG 232
 34 LyLeuTrpAspAlaTrpGlyProTrpSerGluCysSerArgThrCysGly 50
 |||||
 233 GCCATATGGGATGCCCTGGGGCCCATGGAGTGAATGCTCAGCACCTTCGGG 282
 51 GlyGlyAlaSerTyrSerLeuArgArgCysLeuSerSerIysSerCysG 67
 |||||
 283 GGAGGGGCGCTCTACTCTCTGAGCGCTGCTGAGCAGCAAGAGCTGTGA 332
 67 uGlyArgAsnIleArgTyrArgThrCysSerAsnValAspCysProProg 84
 |||||
 333 AGSAGAAATATCCGATACAGAACATGCTGAGTGTGGAGTGTGGCCACAG 382
 84 LuAlaGlyAspPheArgAlaGlnGlnCysSerAlaHisAsnAspVallys 100
 |||||
 383 AGCAGGTGATTTCCGAGCTCAGCAATGCTCAGCTCATATGATGATCAAG 432
 101 HisHisGlyGlnPheTyrGluTrpLeuProValSerAsnAspProAspAs 117
 |||||
 433 CACCATGGCCAGTTTATGAATGGCTTCTGTCTTAATGACCTGCACAA 482
 117 nProCysSerLeuLysCysGlnAlaLysGlyThrThrLeuValValGlu 134
 |||||
 483 CCATGTTTCACTCAAGTGCACGCAAGCAAGCAACCTGTGTGTGAAC 532
 134 euAlaProLysValLeuAspGlyThrArgCystYrThrGluSerLeuAsp 150
 |||||
 533 TAGCACCCTAAGGCTTAGATGGTAGCGTTGCTATACAGAACTTTTGGAT 582
 151 MetCysIleSerGlyLeuCysGlnIleValGlyCysAspHisGlnLeuG 167
 |||||
 583 ATGTGATCATGAGTGTATGCCAAATTTGTGGCTGCCATCACCAGCTGGG 632
 167 ySerThrValLysGluAsnAsnCysGlyValCysAsnGlyAspGlySer 184
 |||||
 633 AAGCACCGTCAAGGAAGATAACTGTGGGGTCTGCAACGGAGATGGGTCCA 682
 184 hrCysArgLeuValArgGlyGlnTyrLysSerGlnLeuSerAlaThrLys 200
 |||||
 683 CCTGCCGGCTGGTCCGAGGGCAGTATAATCCAGCTCTCCGCAACCAAAA 732
 201 SerAspAspThrValValAlaIleProTyrGlySerArgHisIleArgLe 217
 |||||
 733 TCGGATCATAGTGTGGTGTGCACTTCCCTATGGAAGTAGACATATCGCCT 782
 217 uValLeuLysGlyProAspHisLeuTyrLeuGluThrLysThrLeuGlnG 234
 |||||
 783 TGTCTTAAAGGTCCTGATCATCTTATATCTGGAACCAAAACCTCCAGG 832
 234 lyThrLysGlyGluAsnSerLeuSerSerThrGlyThrPheLeuValAsp 250
 |||||
 833 GGACTAAAGGTGAAACACAGTCTCAGCTCCACAGGAACCTTTCCTTGTGGAC 882
 251 AsnSerSerValAspPheGlnLysPheProAspLysGluIleLeuArgme 267
 |||||
 883 AATTCTAGTGTGGACTTCCAGAAATTTCCAGACAAAGAGATACTGAGAAT 932
 267 tAlaGlyProLeuThrAlaAspPheIleValLysIleArgAsnSerGlyS 284
 |||||
 933 GCGTGGACCACTCAGCAGAGATTTCATTTGTCAGAGATTGTAACTCGGGCT 982
 284 erAlaAspSerThrValGlnPheIlePheTyrGlnProIleIleHisArg 300
 |||||
 983 CCGCTGACAGTACAGTCCAGTTCATCTTCTATCAACCCCATCATCCACCGA 1032
 301 TrpArgGluThrAspPhePheProCysSerAlaThrCysGlyGlyClyTy 317
 |||||
 1033 TGGAGGGAGACGGATTTCTTCTTCTGCTCAGCAACCTGTGGAGGAGGTTA 1082
 317 rGlnLeuThrSerAlaGluCysTyrAspLeuArgSerAsnArgValVala 334

```

|||||
1083 TCAGTGCATCGCTGAGTGTAGCACTCTGAGGAGCAACCGTGTGGTGG 1132
334 laAspGlnTyrCysHisTyrTyrProGluAsnIleLysProLysProLys 350
|||||
1133 CTGACCAATAGTGTCTACATATTACCCAGAGAACATCAACACCAACCCCAAG 1182
351 LeuGlnGluCysAsnLeuAspProCysProAlaSerAspGlyTyrLysG1 367
|||||
1183 CTTGAGAGTGCACCTGGATGCTTGTGCTCCAGCCAGTGCAGTACAGCA 1232
367 nileMetProTyrAspLeuTyrHisProLeuProArgTyrGluAlaThrP 384
|||||
1233 GATCATGCTTATACCTCTACATACCTCCCTTCCTCGGTGGAGGCCACCC 1282
384 roTprThrAlaCysSerSerCysGlyGlyGlyIleGlnSerArgAla 400
|||||
1283 CATGACCGGCTGTCTCTCTGCTGCTGGGGGGGCGCATCCAGAGCGGGCA 1332
401 ValSerCysValGluGluAspIleGlnGlyHisValThrSerValGluG1 417
|||||
1333 GTTCTCTGTGGAGGAGGACATCCAGGGGCATGTCTCAGTGAAGA 1382
417 uTrpLysCysMetTyrThrProLysMetProIleAlaGlnProCysAsnI 434
|||||
1383 GTGAAATGCATGTACACCCCTAAGATGCCATCGCGCAGCCCTGCAACA 1432
434 lePheAspCysProLysTyrLeuAlaGlnGluTyrSerProCysThrVal 450
|||||
1433 TTTTTCAGTCCCTAAATGGCTGGCAGCAGTGGTCTCGGTGCACAGTG 1482
451 ThrCysGlyGlnGlyLeuArgTyrArgValValLeuCysIleAspHisAr 467
|||||
1483 ACATGTGGCCAGGGCTCAGATACCGTGTGGTCTCTGCAATCGACCATCG 1532
467 gGlyMetHisThrGlyCysSerProLysThrLysProHisIleLysG 484
|||||
1533 AGGAATGCACACAGGAGGCTGTAGCCCAAAACAAAGCCCATATAAAG 1582
484 luGluCysIleValProThrProCysTyrLysProLysGluLysLeuPro 500
|||||
1583 AGGAATGCATCGTACCCACTCCCTGCTATAAACCAAGAGAACTTCCA 1632
501 ValGluAlaLysLeuProThrPheLysGlnAlaGlnGluLeuGluG1 517
|||||
1633 CTCGAGGCCAAGTGGCATGGTTCAAACAGCTCAAGAGCTAGAGAAGG 1682
517 yAlaAlaValSerGluGluProSerPheIleProGluAla 530
|||||
1683 AGCTGCTGTGTACAGAGGAGCCCTCGTAAGTTGTAAAGCA 1722
seq_name: /SIDS1/gcgdata/hold-geneseq/geneseq - emb1/NA2001A.DAT:AAF63447
seq_documentation_block:
ID AAF63447 standard; cDNA; 1803 BP.
XX AC
XX AAF63447;
XX AC
XX 14-MAY-2001 (first entry)
XX DE
XX Human ADAMTS-R1 cDNA sequence.
XX ADAMTS-N; disintegrin; metalloprotease; thrombospondin type I motif;
KW tumour cachexia; inflammation;; dermatosparaxis; EDS-VIIC; angiogenesis;
KW Ehlers-Danlos syndrome type VIIC; articular cartilage erosion; human;
XX metastasis; embryogenesis; egg implantation; ADAMTS-R1; ss.
XX OS
XX Homo sapiens.
XX PN
XX WO200111074-A2.
XX PD
XX 15-FEB-2001.
XX

```

```

PF 03-AUG-2000; 2000WO-US21223.
XX PR
XX 06-AUG-1999; 99US-0369364.
XX PA
XX (CLEV-) CLEVELAND CLINIC FOUND.
XX PA (APTE/) APTE S S.
XX PA (HURS/) HURSKAINEN T L.
XX PA (HIRO/) HIROHATA S.
XX PI
XX Apte SS, Hurskainen TL, Hirohata S;
XX WPI; 2001-159978/16.
XX DR P-PSDB; AAB72290.
XX PT
XX Murine and human 'A Disintegrin-like And Metalloprotease domain with
XX Thrombospondin type I motifs' proteins and the nucleic acids encoding
XX them, useful for treating e.g. tumours, inflammation and arthritis -
XX Claim 15; Fig 11; 181pp; English.
XX CC
XX This invention relates to murine and human ADAMTS-N (A disintegrin-like
XX and metalloprotease domain with thrombospondin type I motifs) proteins,
XX designated ADAMTS-5, 6, 7, 8, 9, 10 and R1. Also included in the
XX invention are cDNA sequences encoding the proteins, and antibodies
XX specific for the proteins. The nucleic acid sequences and antibodies
XX used in the prevention, diagnosis and treatment of diseases associated
XX with inappropriate ADAMTS-N expression. Disorders that may be treated
XX using the nucleic acids, proteins and antibodies include, for example
XX tumour cachexia, inflammation, dermatosparaxis in cattle or Ehlers-Danlos
XX syndrome type VIIC (EDS-VIIC) in humans, erosion of articular cartilage
XX in arthritic (both inflammatory and non-inflammatory) disease,
XX angiogenesis, tumour growth and metastases, and they may also be used for
XX controlling embryogenesis and implantation of fertilised eggs. The
XX present sequence represents cDNA encoding human ADAMTS-R1.
XX SQ
XX Sequence 1803 BP; 475 A; 460 C; 457 G; 411 T; 0 other;

alignment_scores:
Quality: 2910.00 Length: 559
Ratio: 5.399 Gaps: 1
Percent Similarity: 96.422 Percent Identity: 94.275

alignment_block:
US-10-044-807-2 x AAF63447 ..
Align seg 1/1 to: AAF63447 from: 1 to: 1803
1 MetGluCysCysArgArgAlaThrProGlyThrLeuLeuPheLeuAl 17
|||||
51 ATGGAATGCTGCGTGGGCACTCTGGCACACTGCTCTCTTCTGGC 100
17 aPheLeuLeuLeuSerSerArgThrAlaArgSerGluGluAspArgAspG 34
|||||
101 TTTCTCTGCTCTGAGTCCAGGACCGCACGCTCCGAGGAGGACCGGAGC 150
34 lyLeuTrpAspAlaTrpGlyProTrpSerGluCysSerArgThrCysGly 50
|||||
151 GCCTATGGATGCTGGGGCCCATGGAGTGAATGCTCAGCACCTGCGGG 200
51 GlyGlyAlaSerTyrSerLeuArgArgCysLeuSerSerLysSerCysG1 67
|||||
201 GTTGGGGCCGCAACTCTCTGAGGCGCTGCTGAGCAGCAAGAGCTGTGA 250
67 uGlyArgAsnIleArgTyrArgThrCysSerAsnValAspCysProProG 84
|||||
251 AGGAAGAAATATCCGATACAGAACATGCAGTAATGTGGACTGCCACCAG 300
84 luAlaGlyAspPheArgAlaGlnGlnCysSerAlaHisAsnAspValLys 100
|||||
301 AAGCAGGTGATTTCGAGCTCAGCAATGCTCAGCTCATATGATGTCAG 350
101 HisHisGlyGlnPheTyrGluTrpLeuProValSerAsnAspProAspAs 117

```

```

|||||
351 CACCATGCCAGTATTATGAATGGCTTCTGTGTCTAATGACCCCTGACAA 400
|||||
117 nProCysSerLeuLysCysGlnAlaLysGlyThrThrLeuValValGluL 134
|||||
401 CCCATGTTCACTCAAGTGCAGCAAGCAAGCAAGCAAGCAAGCAAGCAAG 450
|||||
134 euAlaProLysValLeuAspGlyThrArgCysThrThrGluSerLeuAsp 150
|||||
451 TAGCACCTTAAGGTCCTAGATGTCAGCGTGTCTATACAGAACTTCTGGAT 500
|||||
151 MetCysIleSerGlyLeuCysGlnIleValGlyCysAspHisGlnLeuG 167
|||||
501 ATGTGCATCAGTGGTTATGCCAAATGTTGGCTGCGATCACACAGCTGG 550
|||||
167 ySerThrValLysGluAspAsnCysGlyValCysAsnGlyAspGlySerT 184
|||||
551 AAGCACCGTCACAGGAAGATAAATGTTGGGTCTGCAACGGAGATGGGTCCA 600
|||||
184 hrCysArgLeuValArgGlyGlnTyrLysSerGlnLeuSerAlaThrLys 200
|||||
601 CTTGCCCGGTGGTCCGAGGCGAGTATAAATCCAGCTCTCCGCAACCAAA 650
|||||
201 SerAspThrValValAlaIleProTyrGlySerArgHisIleArgLe 217
|||||
651 TCGGATGATACGTGTGCTTCAATTCCTATGGAAGTAGACATATTGCGCT 700
|||||
217 uValLeuLysGlyProAspHisLeuTyrLeuGluThrLysThrLeuGlnG 234
|||||
701 TGTCTTAAAGGTCCTGATCAGTATATCTTGGAAACCAAAACCCCTCAGG 750
|||||
234 lyThrLysGlyGluAsnSerLeuSerSerThrGlyThrPheLeuValAsp 250
|||||
751 GGACTAAAGGTGAACAGCTCTCAGCTCCACAGGAATTTCTCTGTGGAC 800
|||||
251 AsnSerSerValAspPheGlnLysPheProAspLysGluIleLeuArgMe 267
|||||
801 AATTCTAGTGGACTTCCAGAAATTTCCAGCAAAAGAGATACGTAGAAT 850
|||||
267 tAlaGlyProLeuThrAlaAspPheIleValLysIleArgAsnSerGlyS 284
|||||
851 GGTGGACCACTCACAGCAGATTTCATTTCAAGATTCTGATCTCGGGCT 900
|||||
284 erAlaAspSerThrValGlnPheIlePheTyrGlnProIleIleHisArg 300
|||||
901 CCGCTGACAGTACAGTCCAGTTCATCTTCATCAACCCATCATCCACCGA 950
|||||
301 TrpArgGluThrAspPheProCysSerAlaThrCysGlyGlyGlyTy 317
|||||
951 TGGAGGAGACGGATTTCTTTCTTCTGCTGCTCAGCAACCTGTGGAGAGGTTA 1000
|||||
317 rGlnLeuThrSerAlaGluCysTyrAspLeuArgSerAsnArgValValA 334
|||||
1001 TCAGCTGACATCGGCTGAGTGTACGATCTGAGGAGCAACCGTGTGGTTG 1050
|||||
334 laAspGlnTyrCysHisTyrTyrProGluAsnIleLysProLysProLys 350
|||||
1051 CTGACCAATCTGTCACTATTACCCAGAGAACATCAAAACCCCAACCAAG 1100
|||||
351 LeuGlnGluCysAsnLeuAspProCysProAlaSerAspGlyTyrLysG 367
|||||
1101 CTTGAGGAGTGCACCTTGGATCTTGTCTTCCAGCCAGTACGAGTACAGCA 1150
|||||
367 nIleMetProTyrAspLeuTyrHisProLeuProArgTyrGluAlaThrP 384
|||||
1151 GATCATGCCCTTATGACCTCTACCATTCCTTCCCTCGGTGGAGGCCACCC 1200
|||||
384 roTrpThrAlaCysSerSerCysGlyGlyIleGlnSerArgAla 400
|||||
1201 CATGACCGCGTGTCTCTCTCGTGGGGGGGCGCATCCAGAGCGGGCA 1250
|||||
401 ValSerCysValGluGluAspIleGlnGlyHisValThrSerValGluG 417
|||||

```

Sequence	Strd Orig	zScore	EScore	Ien	! Documentation
/cgn2_6/ptodata/2/ina/6A_COMB	seq: US-08-985-526-4	353.00	434.42	1.6e-16	1326 !
/cgn2_6/ptodata/2/ina/6B_COMB	seq: US-08-506-236B-20	317.50	375.65	3.1e-13	3783 !
/cgn2_6/ptodata/2/ina/5B_COMB	seq: US-08-341-843B-1	312.00	368.54	7.7e-13	3774 !
/cgn2_6/ptodata/2/ina/5B_COMB	seq: US-08-427-457E-1	312.00	368.54	7.7e-13	3774 !
/cgn2_6/ptodata/2/ina/5B_COMB	seq: US-08-427-457E-2	312.00	368.54	7.7e-13	3774 !
/cgn2_6/ptodata/2/ina/6B_COMB	seq: US-08-506-236B-13	312.00	368.18	8.1e-13	3888 !
/cgn2_6/ptodata/2/ina/6B_COMB	seq: US-09-041-886-24	311.50	365.47	1.1e-12	4608 !
/cgn2_6/ptodata/2/ina/6B_COMB	seq: PCT-US94-05277-1	305.00	358.92	2.6e-12	3943 !
/cgn2_6/ptodata/2/ina/6B_COMB	seq: US-08-506-236B-27	301.50	349.31	9.1e-12	5993 !
/cgn2_6/ptodata/2/ina/6B_COMB	seq: US-09-383-630-1	301.50	349.31	9.1e-12	5993 !
/cgn2_6/ptodata/2/ina/6B_COMB	seq: US-09-383-630-2	301.50	349.31	9.1e-12	5993 !
/cgn2_6/ptodata/2/ina/5B_COMB	seq: US-08-427-457E-3	287.00	338.14	3.8e-11	3189 !
/cgn2_6/ptodata/2/ina/5A_COMB	seq: US-07-862-021B-11	277.50	325.67	1.9e-10	3226 !
/cgn2_6/ptodata/2/ina/5A_COMB	seq: US-08-313-288B-11	277.50	325.67	1.9e-10	3226 !
/cgn2_6/ptodata/2/ina/5A_COMB	seq: US-08-313-288B-11	277.50	325.67	1.9e-10	3226 !
/cgn2_6/ptodata/2/ina/5A_COMB	seq: PCT-US93-03164-11	277.50	325.67	1.9e-10	3226 !
/cgn2_6/ptodata/2/ina/5A_COMB	seq: US-07-862-021B-9	270.00	313.24	9.2e-10	4029 !
/cgn2_6/ptodata/2/ina/5A_COMB	seq: US-08-313-288B-9	270.00	313.24	9.2e-10	4029 !
/cgn2_6/ptodata/2/ina/6A_COMB	seq: PCT-US93-03164-9	270.00	313.24	9.2e-10	4029 !
/cgn2_6/ptodata/2/ina/6A_COMB	seq: US-08-986-485-1	262.50	301.28	4.3e-09	4843 !
/cgn2_6/ptodata/2/ina/6B_COMB	seq: US-09-335-409-1	262.50	269.14	2.6e-07	68750 !
/cgn2_6/ptodata/2/ina/6B_COMB	seq: US-09-358-402-1	262.50	269.14	2.6e-07	68750 !
/cgn2_6/ptodata/2/ina/6B_COMB	seq: US-09-567-969-1	262.50	269.14	2.6e-07	68750 !
/cgn2_6/ptodata/2/ina/6B_COMB	seq: US-09-568-480-1	262.50	269.14	2.6e-07	68750 !
/cgn2_6/ptodata/2/ina/6B_COMB	seq: US-09-568-486-1	262.50	269.14	2.6e-07	68750 !
/cgn2_6/ptodata/2/ina/6B_COMB	seq: US-09-568-472-1	262.50	269.14	2.6e-07	68750 !
/cgn2_6/ptodata/2/ina/5B_COMB	seq: US-08-447-464-2	261.50	298.03	6.5e-09	5690 !
/cgn2_6/ptodata/2/ina/5B_COMB	seq: US-08-716-679-2	261.50	298.03	6.5e-09	5690 !
/cgn2_6/ptodata/2/ina/5A_COMB	seq: US-08-374-834-17	261.00	306.83	2.1e-09	2610 !
/cgn2_6/ptodata/2/ina/5A_COMB	seq: US-08-644-271-28	261.00	306.83	2.1e-09	2610 !
/cgn2_6/ptodata/2/ina/5A_COMB	seq: US-08-348-006B-6	260.50	296.09	8.3e-09	6000 !
/cgn2_6/ptodata/2/ina/5B_COMB	seq: US-08-800-825A-6	260.50	296.09	8.3e-09	6000 !
/cgn2_6/ptodata/2/ina/6B_COMB	seq: US-09-158-657-6	260.50	296.09	8.3e-09	6000 !
/cgn2_6/ptodata/2/ina/6B_COMB	seq: PCT-US94-10166-6	260.50	296.09	8.3e-09	6000 !
/cgn2_6/ptodata/2/ina/5A_COMB	seq: US-08-245-295-1	260.00	303.89	3.1e-09	2988 !
/cgn2_6/ptodata/2/ina/5A_COMB	seq: US-08-481-130-1	260.00	303.89	3.1e-09	2988 !
/cgn2_6/ptodata/2/ina/5A_COMB	seq: US-08-656-584A-1	260.00	303.89	3.1e-09	2988 !
/cgn2_6/ptodata/2/ina/5A_COMB	seq: US-08-485-604-1	260.00	303.89	3.1e-09	2988 !
/cgn2_6/ptodata/2/ina/5B_COMB	seq: US-08-437-595-1	260.00	303.89	3.1e-09	2988 !
/cgn2_6/ptodata/2/ina/5A_COMB	seq: US-08-474-834-2	259.00	303.08	3.4e-09	2869 !
/cgn2_6/ptodata/2/ina/5A_COMB	seq: US-08-644-271-2	259.00	303.08	3.4e-09	2869 !


```
247 TCTCCA.....TGG..... 255
382 aThrProThrAlaCysSerSerCysGlyGlyGlyIleGlnSerA 399
256 TCCGAGTGGACCTCTGTCTACGAGCTGTGGCAATGGAATTCACAGC 304
399 rAlaValSerCysValGluGluAspIleGlnGlyHisValThrSerVal 415
305 GCGGCGCTCTCGATAGCCTCAACAACCGATGTGAGGCTCTCTCGGTC 354
416 GluGluThrLysCysMetTyrThrProLysMetProIleAlaGlnProCys 432
355 CAG.....ACAGGACCTG 368
432 sAsnIlePheAspCysProLysThrLeuAlaGlnGlu..... 444
369 CCACATTCAGGAGTGTGACAAAGATTTAAACAGAGTGGTGGAGCC 418
445 .....TTPSerProCysThrValThrCysGlyGlnGlyLeuArg 457
419 ACTGGTCCCGTGGTCTATCTGTGTGACATGTGGTGTGATGCTGATC 468
458 TyrArgValValLeuCysIleAspHisArgGlyMetHisThrGly..... 472
469 ACAAGGATCCGGCTCTGCAACTCTCCAGCCCGCCAGATGAATGGGAACC 518
473 .....GlyCysSerProLysThrLysProHisIleLysGluGluCysI 487
519 CTGTGAGGGCAGCGGGGAGACCAAGCTGCCAAGAAAGACGCTGC. 567
487 leValProThrProCysTyrLysProLysGluLysLeuProValGluAla 503
568 .....CCCATCAATGGA 579
504 LysLeu.....ProThrPheLysGlnAlaGlnGluGluGlyAlaAl 519
580 GGCTGGGGTCTTGG..... 594
519 aValSerGluGluProSerPheIleProGluAlaTIPSerAlaCysThrV 536
595 .....TCACCATGGGACATCTGTCTG 616
536 alThrCysGlyValGlyThrGlnValArgIleValArgCysGlnValLeu 552
617 TCACCTGTGGAGGGGTACAGAACT..... 645
553 LeuSerPheSerGlnSerValAlaAspLeuProIleAspGluCysGluG 569
646 .....AGTCTCTCTCTGGTGCAGCTCTAGAAATGACTGAA...GAGAA 683
569 yProLysProAlaSerGln.....ArgAlaCysTyrAlaGlyP 582
684 CAAGAGTGGCCCAATGAGCTGAGGCGGCCCTCCCTATGCTAT..... 726
582 roCysSerGlyGluIleProGluPheAsnProAspGluThrAspGlyLeu 598
726 ..... 726
599 PheGlyGlyLeuGlnAspPheAspGluLeuTyrAspTrpGluTyrGluG 615
727 CACACGAGTTCAG...TAGAANAATACAGGAGTATGCTGTATAG 773
615 yPheThrLysCysSerGluSerCysGlyGlyValGlnGluAlaValV 632
774 CTGCACCTGAGTGTAC.....TGTCAAGAACTCAGTTA 805
632 alSerCysLeuAsnLysGlnThrArgGluProAlaGluGluAsnLeuCys 648
806 CCATCTGC.....AAAAGGTGTCTGCCCATCATGCC...TGC 843
649 ValThrSerArgArgPro.....ProGlnLeuLeuLysSe 660
```

```
844 TCCAATGCCACAGTCTCTGATGGAGATGCTCTCTCGCTGTTGGCCAG 893
660 rCysAsnLeuAspProCysProAlaArgTrpGluIleGlyLysTrpSerP 677
894 CGACTCTGCGCAGCATGGCTGGTCTCCATGG.....TCCGAGTGGACCT 937
677 roCysSerLeuThrCysGlyValGlyLeuGlnThrArgAspValPheCys 693
938 CCGTCTTCTACGAGCTGTGCAATGGAATTCAGCAGCCGCGCTCTCTGTC 987
694 SerHisLeuLeuSerArgGluMetAsnGluThrValIleLeuAlaAspG 710
988 GATAGCCTCAACACCATGTGAGGCTCTCTCGGTC..... 1023
710 uLeuCysArgGlnProLysProSerThrValGlnAlaCysAsn....Argp 726
1024 .....CAGACACGGACCTGCCACATTCAGGAGTGTGACAAAAGAT 1063
726 heAsnCysProProAlaTIP.....TyrProAlaGlnTIPGlnProCysSer 741
1064 TTAACACAGGATGGTGGCTGGAGCCACCTGCTCCCGTGTGCTCATCTGTCT 1113
742 ArgThrCysGlyGlyValGlnLysArgGluValLeuCysLysGln.. 757
1114 GTGACATGTGTGATGTGTGATCACAAGGATCCGGCTCTGCAACTCTCC 1163
758 .....ArgMetAlaAspGlySerPheLeuGluLeuP 768
1164 CAGCCCGCAGATGAATGGAAACCTGTGAAGC.....GAAGCGC 1204
768 roGluThrPheCysSerAlaSerLysProAlaCysGlnGlnAlaCysLys 784
1205 GGGAGACC.....AAAGCCTGCAAG 1224
785 LysAspAspCysPro.....SerGluTIPLeuLeuSerAs 796
1225 AAAGACGCTGCCCATCAATGAGGCTGGGTCTCTTGG.....TCACC 1268
796 pTrpThrGluCysSerThrSerCysGlyGluGlyThrGlnThrArgSerA 813
1269 ATGGACATCTGTCTGTCTACCTGTGGAGGAGGGGTACAGAAAGCTAGTC 1318
813 laIle 814
1319 GTCCTC 1323
seq_name: /cgn2_5/ptodata/2/ina/6B_seq.us-08-506-296B-20
seq_documentation_block:
; Sequence 20, Application US/08506296B
; Patent No. 6313265
; GENERAL INFORMATION:
; APPLICANT: Phillips, Greg
; APPLICANT: Cunningham, Bruce A.
; APPLICANT: Crossin, Kathryn L.
; TITLE OF INVENTION: NEURITE OUTGROWTH-PROMOTING POLYPEPTIDES
; TITLE OF INVENTION: CONTAINING FIBRONECTIN TYPE III REPEATS AND METHODS OF USE
; NUMBER OF SEQUENCES: 77
; CORRESPONDENCE ADDRESS:
; ADDRESS: The Scripps Research Institute
; STREET: 10550 NO. 6313265th Torrey Pines Road, TPC-8
; CITY: La Jolla
; STATE: California
; COUNTRY: U.S.
; ZIP: 92037
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/506,296B
; FILING DATE: 24-JUL-1995
```

```

/ CLASSIFICATION: 514
/ ATTORNEY/AGENT INFORMATION:
/ NAME: Fitting, Thomas
/ REGISTRATION NUMBER: 34,163
/ REFERENCE/DOCKET NUMBER: TSRI 488.0
/ TELECOMMUNICATION INFORMATION:
/ TELEPHONE: (619) 554-2937
/ TELEFAX: (619) 554-6312
/ INFORMATION FOR SEQ ID NO: 20:
/ SEQUENCE CHARACTERISTICS:
/ LENGTH: 3783 base pairs
/ TYPE: nucleic acid
/ STRANDEDNESS: double
/ TOPOLOGY: linear
/ MOLECULE TYPE: cDNA
/ HYPOTHETICAL: NO
/ ANTI-SENSE: NO
/ FEATURE:
/ NAME/KEY: CDS
/ LOCATION: 1..3783
/ US-08-506-296B-20

```

```

alignment_scores:
  Quality: 317.50      Length: 1121
  Ratio: 0.692        Gaps: 54
  Percent Similarity: 40.946      Percent Identity: 20.607

alignment_block:
US-10-044-807-2 x US-08-506-296B-20 ..
Align seg 1/1 to: US-08-506-296B-20 from: 1 to: 3783

834 ProLeuProPheSerSerSeriLeArg.....ProCysMe 845
|||||
45 CCCCTGCTGCTATACAGATCCAGACGAATATAAGGACACATGTC 94
845 tLeuAlaThrCysAlaArgProGlyArgProSer.ThrLysHisSerPro 861
|||||
95 TAGAGCCACCTGT.....CATCAGGAACAGTCATCCA 126
862 HisIleAlaAlaAlaArgLysValTyrIleGlnThrArgGlnArgLys 878
|||||
127 .....CGCG 131
878 sLeuHisPheValValGlyGlyPheAlaTyrLeuLeuProLysThrAlav 895
|||||
132 CCG.....GTGCTTCCCAACAGATGACA 157
895 alValLeuArgCysProAlaArgArgValArgLysProLeuIleThrTrp 911
|||||
158 TAAGCTGAATGTGAAGCCAGAGGAGACCCCAAGTGGAGTTCCGCTGG 207
912 GluLysAspGlyGlnHisLeuIleSerSerThrHisValThrVal..... 926
|||||
208 ACCAAGATGGCATCCACTCAACCCAGAGAGAAATGGGTGTAGTGGT 257
927 .....AlaProPhe...GlyTyrLeuLysIle.....HisArgL 937
|||||
258 GCATGAGGACCCCTATTCTGGCTCTTCCATCCAGGAGCAACACAGCT 307
937 euLysProSerAspAlaGlyValTyrThrCysSerAla..... 949
|||||
308 TTGCCAGAGGTTTCAGGCACTATCTGCTGCTATGCCAGCAATAGCTA 357
950 GlyProAlaArgGluHisPheValIleLysLeuIleGlyCysAsnArgLys 966
|||||
358 GGAACCTGCCATGTCGATGATC.....CA 383
966 sLeuValAlaAlaArgProLeuSerProArgSerGluGluValLeuAlaG 983
|||||
384 GCTCGTGGCC..... 393

```

```

983 lYArgLysGlyGlyProLysGluAlaLeuGlnThrHisLysHisGlnAsn 999
|||||
394 .....GAGGGTCCCCCAG..... 408
1000 GlyIlePheSerAsnGlySerLysAlaGluLysArgGlyLeuAlaAlaAs 1016
408 ..... 408
1016 nProGlySerArgTyrAspAspLeuValSerArgLeuLeuGlnGlyG 1033
408 ..... 408
1033 lYTrpProGlyLeuLeuAlaSerTrpGluAlaGlnAspSerAlaGlu 1049
|||||
409 ..TGCCCGAAGGAGACTGTAAACCTGTGAAGTGGAGGAGGAGAAATCA 456
1050 ArgAsnThrThrSerSerGluGluAspProGlyAlaGluGlnValLeu 1066
457 GTAGTTCTGCTTGCACACCTCCACCCAGTGCA..... 489
1066 sLeuProPheThrMetValThrGluGlnArgArgLeuAspIleLeuG 1083
|||||
490 ..GCCCCACCTAGGATCTACTGGATGACACAGCAAGATTTTCGACATC 534
1083 lYAsnLeuSerGlnGlnProGluGluLeuArg.....Asp 1094
|||||
535 .....AACAAGATGAGCGGGTGTCCATGGGCCAGANTGGAGAC 573
1095 LeuTyrSerLysHisLeuValAla.....GlnLeuAlaGlnGluIlePh 1109
|||||
574 CTATATTGTGCCAATGTCTTACCTCAGACAATCATTCAGACTACATCTG 623
1109 eArgSerHisLeuGluHisGlnAspThrLeuLeu.....LysProSerG 1124
624 CAATGCCCACTTCCCTGTACCCGACCATCATTCMAAAGGAACCTATTG 673
1124 luArgArgThrSerProValThrLeuSerProHisLysHisValSerGly 1140
674 ACCTCCGGGTCAAGCC..... 690
1141 PheSerSerSerLeuArgThrSerSerThrGlyAspAlaGlyGlyGly 1157
691 .....ACCAACAGCATGATTGAC..... 708
1157 rArgArgProHisArgLysProThrIleLeuArgLysIleSerAlaAla 1174
709 .....CGAAGCCACGCTGCTCTTTCCACCAAC..... 738
1174 lnGlnLeuSerAlaSerGluValValThrHisLeuGlyGlnThrVala 1190
739 .....TCCAGCAGCGCGCTGTAGCTTGCAGGGCCAG..... 771
1191 LeuAlaSerGlyThrLeuSerValLeuLeuHisCysGluAlaIleGly 1207
772 .....TCATTGATCTGGAGTGCAATGCTGAGGCAAT 803
1207 sProArgProThrIleSerTrpAlaArgAsnGlyGluGluValGlnPhe 1224
804 CCCTACACCCACCATCAAGTGGCTGCACCCAGTGCACCAATGCCA...A 850
1224 eAspArgIleLeuLeuGln...ProAspAspSerLeuGlnIleLeuAla 1239
851 CAGACCGTGTATCTACCAAAACCAACAGACCCCTGCAACTACTCAAT 900
1240 ProValGluAlaAspValGlyPheTyrThrCysAsnAlaThrAsnAla 1256
901 GTGGCGGAGAGACGACGATGGCGAGTATACCTGCTTCTGAGAACTCGCT 950
1256 uGlyTyrAspSerValSerIleAlaValThrLeuAlaGlyLysProLeu 1273
951 GGCAGTGGCCGCGCATGCTTACTGATGAGCTGCGCCCATATT 1000
1273 allYThrSerArgMetThrValIleAsnThrGluLysProAlaValThr 1289

```



```

1724 oCysGluAsnMetGluCysArgAspThrThrArg...TyrCysGluLysV 1740
    : : : : : ||||| ||| ||| ||| ||| : : :
2731 AGACCCAGAGGGAGTGGCTGGCCACCTGAGGATTACACTGGAGTG 2780
    : : : : : ||||| ||| ||| ||| ||| : : :
1740 allYsGlnLeuLysLeuCysGlnLeuSerGlnPheLysSerArgCysCys 1756
    : : : : : ||||| ||| ||| ||| ||| : : :
2781 TCAGTCGGACACTAGTCGTCTAC.....TGCAC 2809
    : : : : : ||||| ||| ||| ||| ||| : : :
1757 GlyThr 1758
    : : : : : ||||| ||| ||| ||| ||| : : :
2810 GGCAGC 2815
    : : : : : ||||| ||| ||| ||| ||| : : :

```

```
seq_name: /cgn2_6/ptodata/2/ina/5B_COMB.seq.us-08-341-843B-1
```

```

seq_documentation_block:
; Sequence 1, Application US/08341843B
; Patent No. 5872225
; GENERAL INFORMATION:
; APPLICANT: Lemmon, Vance
; TITLE OF INVENTION: A Method for Characterizing the
; TITLE OF INVENTION: Nucleotide Sequence of LICAM and
; Patent No. 5872225
; TITLE OF INVENTION: the Nucleotide Sequence
; TITLE OF INVENTION: Characterized Thereby
; NUMBER OF SEQUENCES: 39
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fay, Sharpe, Beall, Pagan,
; ADDRESSEE: Minnich & McKee
; STREET: 1100 Superior Avenue
; STREET: Suite 700
; CITY: Cleveland
; STATE: Ohio
; COUNTRY: U.S.A.
; ZIP: 44114-2518
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 720 Kb
; MEDIUM TYPE: storable
; COMPUTER: Compaq Prolinea 5100e
; OPERATING SYSTEM: DOS 5.0
; SOFTWARE: ASCII
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: 07/904,991
; FILING DATE: June 26, 1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Minnich, Richard J.
; REGISTRATION NUMBER: 24,175
; REFERENCE/DOCKET NUMBER: CWR 2 149-1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (216) 861-5582
; TELEFAX: (216) 241-1666
; TELEX: (216) 980162
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3774
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; HYPOTHETICAL: Irrelevant
; ANTI-SENSE: no
; ORIGINAL SOURCE:
; ORGANISM: Homo Sapiens
; INDIVIDUAL ISOLATE: 17-18 week fetus
; IMMEDIATE SOURCE:
; LIBRARY: Stratagene cDNA Library 936206
; CLONE: synthesis of 4 clones
; PUBLICATION INFORMATION:
; AUTHORS: Hlavin, Mary Louise

```

```

; AUTHORS: Lemmon, Vance
; TITLE: Molecular structure and functional
; TITLE: testing of human LICAM: an
; TITLE: interspecies comparison..
; JOURNAL: GENOMICS
; VOLUME: 11
; ISSUE:
; PAGES: 416-423
; DATE: 1991
; RELEVANT RESIDUES IN SEQ ID NO: 1 to 3774
; US-08-341-843B-1

alignment_scores:
    Quality: 312.00      Length: 815
    Ratio: 0.904        Gaps: 30
    Percent Similarity: 42.331    Percent Identity: 20.368

alignment_block:
US-10-044-807-2 x US-08-341-843B-1 ..
Align seg 1/1 to: US-08-341-843B-1 from: 1 to: 3774

843 ProCysMetLeuAlaThrCysAlaArgProGlyArgProSerThrLysHi 859
|||||:|||||
46 CCTGCTGCTTATCCAGATC.....CCGAGGAATATGA 80
859 sserProHisIleAlaAlaAlaArgLysValTyrIleGlnThrArgArg 876
|||||:|||||
81 AGCACACCATGTGATGGAGCCACCTGTCATCAGCAAGACAGTCTCCACGC 130
|||||:|||||
876 InArgLysLeuHisPheValValGlyGlyPheAlaTyrLeuLeuProLys 892
||:|||||
131 GC.....CTGGTTGCTCTCCCA 150
893 ThrAlaValValLeuArgCysProAlaArgValArgLysProLeuIl 909
|||:|||||
151 GATGACATCAGCCTCAAGTGTGAGGCCAGTGCAGCCGCAAGTGCAGTT 200
909 eThrTripGluLysAspGlyGlnHisLeuIleSerSerThrHisValThrV 926
|||:|||||
201 CCGTGGAGGAGGATGTGTCCTCACTTCAAAACCCCAAGGAGAGTGGGTG 250
|||:|||||
926 alAlaProPheGlyTyrLeuLysIleHisArgLeuLysProSerAspAla 942
|||:|||||
251 TGACCGGTGTAC.....CAGTCGCCCCACTCT 276
943 GlyValTyrThrCysSerAlaGlyProAlaArgGluHisPheValIleLy 959
|||:|||||
277 GGCTCCTTCACCATCAGGGCAAC.....AACAGCAACTTGTCTCAGAG 320
959 sLeuIleGlyGlyAsnArgLysLeuValAlaArgProLeuSerProArgS 976
|||:|||||
321 GTTCCAGGGCATCTACCGCTGTTCAGCAATAAGCTGGGCGCCGCCA 370
976 erGluGluGluValLeuAlaGlyArgLysGlyGlyProLysGluAlaIeu 992
|||:|||||
371 TGTCCTATGATCCGGCTCATGGCCGAGGGTGCCCCCAG.....411
993 GlnThrHisLysHisGlnAsnGlyIlePheSerAsnGlySerLysAlaGl 1009
411 .....
1009 uLysArgGlyLeuAlaAlaAsnProGlySerArgTyrAspLeuValS 1026
411 .....
1026 erArgLeuLeuGluGlnGlyTyrTripproGlyGluLeuLeuAlaSerTrip 1042
|||:|||||
412 .....TGGCCAAAGGAGACAGTGAAGCCCGTG 438
1043 GluAlaGlnAspSerAlaGluArgAsnThrThrSerGluGluAspProGl 1059
|||:|||||

```

1341	rAsnValSerSerSerAspGlnGlyLeuTyfSerCysArgAlaAAsnL	1358
:	: : : : : : :	:
1176	CAACGGTCAGCCAGTGCACAATGGTGACCCTAAATGTAGGCCCAACC	1225
:	: : : : : : :	:
1358	eufHisGlyLuLeuSerThrGlnLeuLeulleAspPro	1374
:	: : : : : : :	:
1226	GGCACGGCGCTTTGCTGGCCAATGCCATCATCTACGTTGTC.....	1266
:	: : : : : : :	:
1375	GlnValProThrGlnLeuGluAspIleArgAlaLeuLeuAlaAThrGl	1391
:	: : : : : : :	:
1267	CAGCTGCCGCCAAG.....	1281
:	: : : : : : :	:
1391	yProAsnLeuProSerValLeuThrSerProLeuGlyThrGlnLeuValL	1408
:	: : : : : : :	:
1282ATCTTGACTCGCGCAATCAGACGTACATGGCGTG	1315
:	: : : : : : :	:
1408	euAspProGlyAsnSerAlaLeuLeuGlyCysProIleLeuGlyHisPro	1424
:	: : : : : : :	:
1316	TCCAG...GGCAGCAGTCGCTTCTGTGCAGGCCCTTCGAGCGCCT	1362
:	: : : : : : :	:
1425	ValProAsnIleThrTrpPhe...HisGlyGlnProIleValThrAl	1440
:	: : : : : : :	:
1363	GTGCCAGTGTTCAGTGGCTGGACGAGGATGGCAACAGTAGCTTCAGGA	1412
:	: : : : : : :	:
1440	aThrGlyLeuThrHisHisIleLeuAlaAlaGlyGlnIleLeuGlnValA	1457
:	: : : : : : :	:
1413	CGAACGGTCTTCCTCCCTAT.....GCCAATGGGACCTGGCGATTTC	1453
:	: : : : : : :	:
1457	laAsnLeuSerGlyGlySerGlnGlyGluPheSerCysLeuAlaGlnAsn	1473
:	: : : : : : :	:
1454	GAGACCTCCAGGCCAATGACACCGGACGCTACTTCTGCCTGGCTGCCAAT	1503
:	: : : : : : :	:
1474	GluAlaGlyValLeuMetGlnIlysAlaSerLeuValIleGlnAspTryr	1490
:	: : : : : : :	:
1504	GACCAAAACAATGTTACCATCATGCTACCTCAAGGTTAAAGAT....	1548
:	: : : : : : :	:
1490	pTripSerValAspArgLeuAlaThrCysSerAlaSerCysGlyAsnArg	1507
:	: : : : : : :	:
1549GCAACTCAGA	1558
:	: : : : : : :	:
1507	lyValGlnGlnProArgLeuArgCysLeuLeuAsnSerThrGluValAsn	1523
:	: : : : : : :	:
1559	TCACCTAGGGCCCCGCGACACAAATCGAAGAAAAGGTTCCAGGGTGACC	1608
:	: : : : : : :	:
1524	ProAlaHisCysAlaGlyLysValArgProAlaValGlnProfileAlaCy	1540
:	: : : : : : :	:
1609	..TTCAGTGCAGCGCTCCCTTTGNACCCTCTCTTCGAGCCCAGCATCAC	1655
:	: : : : : : :	:
1540	sAsnArgArgAspCys.....	1545
:	: : : : : : :	:
1656	CTGGCTGGGACGCTCGAGACCTCCAGGAGCTTGGGGACAGTGACAAGT	1705
:	: : : : : : :	:
1546ProSerArg.TrpMetValThrSerTrpSerAlaCysTh	1558
:	: : : : : : :	:
1706	ACTTCATAGAGATGGCGCGCTGGTCATCCACAGCTGGAGTACAGCGAC	1755
:	: : : : : : :	:
1558	rArgSerCysGlyGlyValGlnThrArgArgValThrCysGlnIlysL	1575
:	: : : : : : :	:
1756	CAGG.....	1759
:	: : : : : : :	:
1575	euLysAlaSerGlyIleSerThrProValSerAsnAspMetCysThrGln	1591
:	: : : : : : :	:
1760GCAACTACAGCTGGTGGCCAGTACCGAACTGGATGTGGTGAGA	1804
:	: : : : : : :	:
1592	ValAlaLysArgProValAspThrGlnAlaCysAsnGlnGlnLeuCysVa	1608
:	: : : : : : :	:
1805	GTAGGCGACAGC.....	1816
:	: : : : : : :	:
1608	IcLutrpAlaPheSerSerTrpGlyGlnCysAsnGlyProCys	1622
:	: : : : : : :	:
1817	.TCITGGTGGGAGCCCTGGCGGGTGCACCGCTGGTGC	1858
:	: : : : : : :	:

seq_name: /cgn2_6/ptodata/2/ina/5B_COMB.seq:US-08-427-497E-1

seq_documentation_block:
; Sequence 1, Application US/08427497E
; Patent No. 5969124
; GENERAL INFORMATION:
; APPLICANT: Lemmon, Vance
; TITLE OF INVENTION: A Method for Characterizing the
; TITLE OF INVENTION: Nucleotide Sequence of L1CAM and
; Patent No. 5969124
; TITLE OF INVENTION: the Nucleotide Sequence
; TITLE OF INVENTION: Characterized Thereby
; NUMBER OF SEQUENCES: 44
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fay, Sharpe, Beall, Fagan,
; ADDRESSEE: Minnich & McKee
; STREET: 1100 Superior Avenue
; STREET: Suite 700
; CITY: Cleveland
; STATE: Ohio
; COUNTRY: U.S.A.
; ZIP: 44114-2518
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 720 kb
; MEDIUM TYPE: Storable
; COMPUTER: Compaq Prolinea 5100e
; OPERATING SYSTEM: DOS 5.0
; SOFTWARE: ASCII
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: 07/904,991
; FILING DATE: June 26, 1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Minnich, Richard J.
; REGISTRATION NUMBER: 24,175
; REFERENCE/DOCKET NUMBER: CWR 2 149-3-1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (216) 861-5582
; TELEFAX: (216) 241-1666
; TELEX: (216) 980162
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3774
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: nucleic acids
; HYPOTHETICAL: Irrelevant
; ANTI-SENSE: no
; ORIGINAL SOURCE:
; ORGANISM: Homo Sapiens
; INDIVIDUAL ISOLATE: 17-18 week fetus
; IMMEDIATE SOURCE:
; LIBRARY: Stratagene cDNA Library 936206
; CLONE: synthesis of 4 clones
; PUBLICATION INFORMATION:
; AUTHORS: Hlavin, Mary Louise
; AUTHORS: Lemmon, Vance
; TITLE: Molecular structure and functional
; TITLE: testing of human L1CAM: an
; TITLE: interspecies comparison.
; JOURNAL: GENOMICS
; VOLUME: 11
; ISSUE:
; PAGES: 416-423
; DATE: 1991
; RELEVANT RESIDUES IN SEQ ID NO: 1 to 3774

alignment_scores:
Quality: 312.00 Length: 815
Ratio: 0.904 Gaps: 30
Percent Similarity: 42.331 Percent Identity: 20.368

alignment_block:
US-10-044-807-2 x US-08-427-497E-1 ..
Align seg 1/1 to: US-08-427-497E-1 from: 1 to: 3774

843 ProCysMetLeuAlaThrCysAlaArgProGlyArgProSerThrLysHi 859
|||||:|||||
46 CCTGCTGCTTATCCAGATC.....CCGAGGAATATCA 80
859 sSerProHisIleAlaAlaAlaArgLysValTyrIleGlnThrArgArg 876
:|||||:|||||
81 AGGACACCATGTGATGGAGCCACCTGTCATCACGGAACAGTCTCCACGGC 130
:|||||:|||||
876 InArgLysLeuHisPheValValGlyGlyPheAlaTyrLeuLeuProLys 892
:|||||:|||||
131 GC.....CTGGTGTCTTCCCCACA 150
893 ThrAlaValValLeuArgCysProAlaArgValArgLysProLeuI 909
:|||||:|||||
151 GATGACATCAGCTCAAGTGTGAGGCCAGTGCAGCCGAGAGTCCAGTT 200
909 eThrTrpGluLysaspGlyGlnHisLeuIleSerSerThrHisValThrV 926
:|||||:|||||
201 CGCTGGACGAGGATGTGCTCCACTTCAAAACCCCAAGGAGAGTGGTGTG 250
:|||||:|||||
926 alAlaProPheGlyTyrLeuLysIleHisArgLeuLysProSerAspAla 942
:|||||:|||||
251 TGACGGTGTAC.....CAGTGGCCCCACTCT 276
943 GlyValTyrThrCysSerAlaGlyProAlaArgGluHisPheValIle 959
:|||||:|||||
277 GGCTCCTTCACCATCAGGCAAC.....AACACCAACTTGTCTCAGAG 320
959 sLeuIleGlyGlyAsnArgLysLeuValAlaArgProLeuSerProArg 976
:|||||:|||||
321 GTTCCAGGGCATCTACCGCTGCTTTGCCAGCAATAGCTGGCACCGCCA 370
976 erGluGluGluValLeuAlaGlyArgLysGlyGlyProLysGluAlaLeu 992
:|||||:|||||
371 TGTCCTAGATCGGCTCATGCCGAGGTGCCCAAG.....411
993 GlnThrHisLysHisGlnAsnGlyIlePheSerAsnGlySerLysAlaGl 1009
411
1009 uLysArgGlyLeuAlaAlaAsnProGlySerArgTyrAspAspLeuVal 1026
411
1026 erArgLeuLeuGluGlnGlyTyrProGlyGluLeuLeuAlaSerTrp 1042
|||||:|||||
412TGGCCAAAGGAGACAGTGAAGCCCGTG 438
1043 GluAlaGlnAspSerAlaGluArgAsnThrThrserGluGluAspProGl 1059
:|||||:|||||
439 GAGGTGGAGGAGGGAGTCACTGCTGCTTGCCTTGCACACCTCCCCCAAG 488
1059 yAlaGluGlnVal.....LeuLeuHisLeuP 1068
:|||||:|||||
489 TGCAGAGCCCTCTCCGGATCTACTGGATGAACAGCAAGATCTTCACATCA 538
1068 roPheThrMetValThrGluGlnArgLeuAspAspIleLeuGlyAsn 1084
:|||||:|||||
539 AG.....CAGGACGAGGGGTGACGATGGGCCAGAACGGCAAC 576
1085 LeuSerGlnGlnProGluGluLeuArgAspLeuTyrSerLysHisLeuVa 1101
|||:|||||
577 CTCTACTTTGCCAATGTGCTCACCTCCGACACCACTCAGACTAC.....621

Wed Jul 24 11:35:44 2002

```
1101 lAlaGlnLeuAlaGlnGluIlePheArgSerHisLeuGluHisGlnAspT 1118
    ||| : : : : : ||| :
622 .....ATCTGCCACGCCACATCTCCAGGACCACAGGA 652
1118 hrLeuLeuLysProSerGluArgArgThrSerProValThrLeuSerPro 1134
    ||| : : : : : ||| :
653 CCATCATT.....CAGAAGGAACCCATT..... 675
1135 HisLysHisValSerGlyPheSerSerSerLeuArgThrSerSerThrG1 1151
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
676 .....GACCTCCGGGTCAAGGCCACCC... 696
1151 yAspAlaGlyGlyGlySerArgArgProHisArgLysProThrIleLeuA 1168
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
697 .....AACAGCATGATGACAGGAGCCGCGCTGCTCT 730
1168 rglYsIleSerAlaAlaGlnGlnLeuSerAlaSerGluValValThrHis 1184
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
731 TCCCCACCAAC.....TCCAGCAGCCACCTGGTGGCCTTG 765
1185 LeuGlyGlnThrValAlaLeuAlaSerGlyThrLeuSerValLeuLeuHi 1201
    ||||| : : : : : : : : : : : : : : : : : : : : : : : :
766 CAGGGGCAG.....CCATTGGTCCCTGGA 788
1201 sCysGluAlaIleGlyHisProArgProThrIleSerThrAlaArgAsnG 1218
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
789 GTGATCGCCGAGGCTTCTCCAGGCCACCAATCAATGCTGGCCCCA 838
1218 lyGluGluValGlnPheSerAspArgIleLeuLeuGln...ProAspAsp 1233
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
839 GTGGCCCCATGCCA...GCTGACCGCTGCACCTACCAAGAACCAACAAG 885
1234 SerLeuGlnIleLeuAlaProValGluAlaAspValGlyPheThrCy 1250
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
886 ACCGTGACGTGCTGAAGTGGGAGGAGGATGATGCGGAGTACCGCTG 935
1250 sAsnAlaThrAsnAlaLeuGlyTyArgSerValSerIleAlaValThrL 1267
    | ||| ||| : : : : : ||| : : : : :
936 CTGCGCGGAGAACTCACTGGCAGTGGCCGCGCATCGTACTATGTCACG 985
1267 euAlaGlyLysProLeuValLysThrSerArgMetThrValIleAsnThr 1283
    : : : : : ||| : : : : : : : : : : : : : : : : : : : :
986 TGGAGGCTGCCCCCTACTGG.....CTG 1008
1284 GluLysProAlaValThrValAspIleGlySerThrIleLysThrValG1 1300
    : : : : : ||||| : : : : : : : : : : : : : : : : : : :
1009 CACAAGCCC.....CAGAGCCATCTATATGGGCC 1037
1300 nGlyValAsnValThrIleAsnCysGlnValAlaGlyValProGluAlaG 1317
    ||| : : : : : : : : : : : : : : : : : : : : : : : : : :
1038 AGGAGAGACTCCCGCTGGACTGCCAAGTCCAGGGCAGGCCCAACCAAG 1087
1317 luValThrTrpPheArgAsnLysSerLysLeuGlySerPro..... 1330
    ||||| ||||| ||| : : : : :
1088 AGGTACACCTGGAGAATCAAC.....GGGATCCCTGTGGAGAG 1125
1331 .....HisHisLeuHisGluGlySerLeuLeuLeuTh 1341
1126 CTGGCCAAGACCAGAACTACCGGATTACCGTGGCGCCCTGATCCTCAG 1175
1341 rAsnValSerSerAspGlnGlyLeuTyTrpSerCysArgAlaAlaAsnL 1358
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1176 CAACGTGACGCCAGTACACAAATGTTGACCCCAATGTGAGGCCGCCAACC 1225
1358 euHisGlyCluLeuThrGluSerThrGlnLeuLeuIleLeuAspPro 1374
    ||||| ||| : : : : : : : : : : : : : : : : : : : :
1226 GCGAGGGCTCTGTGGCCAAATGCTACATCTACGTGTC..... 1266
1375 GlnValProThrGlnGluAspIleArgAlaLeuLeuAlaAlaThrG1 1391
    ||| : : : : : : : : : : : : : : : : : : : : : : : :
1267 CAGCTGCCAGCCCAAG..... 1281
```

```
seq_name: /cgn_6/ptodata/2/ina/5B_COMB.seq:US-08-427-497E-2
seq_documentation_block:
; Sequence 2, Application US/08427497E
; Patent No. 5969124
; GENERAL INFORMATION:
; APPLICANT: Lemmon, Vance
; TITLE OF INVENTION: A Method for Characterizing the
; TITLE OF INVENTION: Nucleotide Sequence of L1CAM and
; Patent No. 5969124
; TITLE OF INVENTION: the Nucleotide Sequence
; TITLE OF INVENTION: Characterized thereby
; NUMBER OF SEQUENCES: 44
```


; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Fay, Sharpe, Beall, Fagan,
 ; ADDRESSEE: Minnich & McKee
 ; STREET: 1100 Superior Avenue
 ; CITY: Cleveland
 ; STATE: Ohio
 ; COUNTRY: U.S.A.
 ; ZIP: 44114-2518
 ;
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Diskette, 3.50 inch, 720 kb
 ; MEDIUM TYPE: storable
 ; COMPUTER: Compaq Prolinea 5100e
 ; OPERATING SYSTEM: DOS 5.0
 ; SOFTWARE: ASCII
 ;
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: 07/904,991
 ; FILING DATE: June 26, 1992
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Minnich, Richard J.
 ; REGISTRATION NUMBER: 24,175
 ; REFERENCE/DOCKET NUMBER: CWR 2 149-3-1
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: (216) 861-5582
 ; TELEFAX: (216) 241-1666
 ; TELEX: (216) 980162
 ; INFORMATION FOR SEQ ID NO: 2:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 3774
 ; TYPE: nucleic acid
 ; STRANDEDNESS: single
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: nucleic acids
 ; HYPOTHETICAL: irrelevant
 ; ANTI-SENSE: no
 ; ORIGINAL SOURCE:
 ; ORGANISM: Homo Sapiens
 ; INDIVIDUAL ISOLATE: 17-18 week fetus
 ; IMMEDIATE SOURCE:
 ; LIBRARY: Stratagene cDNA Library 936206
 ; CLONE: synthesis of 4 clones
 ; PUBLICATION INFORMATION:
 ; AUTHORS: Hlavin, Mary Louise
 ; AUTHORS: Lemmon, Vance
 ; TITLE: Molecular structure and functional testing of
 ; TITLE: human L1CAM: an interspecies comparison.
 ; JOURNAL: GENOMICS
 ; VOLUME: 11
 ; ISSUE:
 ; PAGES: 416-423
 ; DATE: 1991
 ; RELEVANT RESIDUES IN SEQ ID NO: 1 to 3774
 ; US-08-427-497E-2

alignment_scores:
 Quality: 312.00 Length: 815
 Ratio: 0.904 Gaps: 30
 Percent Similarity: 42.331 Percent Identity: 20.368
 alignment_block:
 US-10-044-807-2 x US-08-427-497E-2 ..
 Align seg 1/1 to: US-08-427-497E-2 from: 1 to: 3774

843 ProCysMetLeuAlaThrCysAlaArgProGlyArgProSerThrLysHI 859
 |||||
 46 CCCTGCTGCTATCCAGATC.....CCCGAGGAATATGA 80

859 sserProHisIleAlaAlaAlaArgLysValTyrIleGlnThrArgArg 876
 : : : : :
 81 AGGACACCATGTGATGGAGCCACCTGTCTACGGAACAGCTCTCCACGGC 130
 : : : : :
 876 lnArgLysLeuHisPheValValGlyGlyPheAlaTyrLeuLeuProLys 892
 : : : : :
 131 GC.....CTGGTTGTCTTCCCCACA 150
 : : : : :
 893 ThrAlaValValLeuArgCysProAlaArgValArgLysProLeuI 909
 : : : : :
 151 GATGACATCAGCTCAAGTGTGAGGCCAGTGGCAAGCCCGAGTCAGTT 200
 : : : : :
 909 eThrTrpGluLysAspGlyGlnHisLeuIleSerSerThrHisValThrV 926
 : : : : :
 201 CCGCTGGACGAGGGATGCTCTCAAAACCCCAAGGAAGAGCTGGGTG 250
 : : : : :
 926 alAlaProPheGlyTyrLeuLysIleHisArgLeuLysProSerAspAla 942
 : : : : :
 251 TGACCGGTGTAC.....CAGTCGCCCACTCT 276
 : : : : :
 943 GlyValTyrThrCysSerAlaGlyProAlaArgGluHisPheValIleL 959
 : : : : :
 277 GGTCTCTTCAACCATCACGGGCAAC.....AACAGAACTTGTCTCAGAG 320
 : : : : :
 959 sLeuIleGlyGlyAsnArgLysLeuValAlaArgProLeuSerProArgS 976
 : : : : :
 321 GTTCCAGGGCATCTACCGCTCTTCCAGCAATAAGCTGGGACCCGCCA 370
 : : : : :
 976 erGluGluGluValLeuAlaGlyArgLysGlyGlyProLysGluAlaLeu 992
 : : : : :
 371 TGTCCTCATGATCCGGCTCATGGCGAGGGTGGCCCAAG.....411
 : : : : :
 993 GlnThrHisLysHisGlnAsnGlyIlePheSerAsnGlySerLysAlaGl 1009
 : : : : :
 411 411
 1009 uLysArgGlyLeuAlaAlaAsnProGlySerArgTyrAspLeuValS 1026
 : : : : :
 411 411
 1026 erArgLeuLeuGluGlnGlyTrpProGlyGluLeuLeuAlaSerTrp 1042
 : : : : :
 412TGGCCAAAGAGAGACAGTGAAGCCCGTG 438
 : : : : :
 1043 GluAlaGlnAspSerAlaGluArgAsnThrThrSerGluGluAspProGl 1059
 : : : : :
 439 GAGGTGGAGGAAGGGAGTCACTGCTGCTTCCCTTGCCTTGCACCACTCC 488
 : : : : :
 1059 yAlaGluGlnVal.....LeuLeuHisLeup 1068
 : : : : :
 489 TGCAGAGCCTCTCCGGATCTACTGGATGAACAGCAAGATCTTGCACATCA 538
 : : : : :
 1068 roPheThrMetValThrGluGlnArgArgLeuAspIleLeuGlyAsn 1084
 : : : : :
 539 AG.....CAGGACGAGCGGTGACGATGGCGCCAGACGCGCAAC 576
 : : : : :
 1085 LeuSerGlnGlnProGluGluLeuArgAspLeuTyrSerLysHisLeuVa 1101
 : : : : :
 577 CTCTACTTGTCCCAATGTGCTCACCCTCCGACCAACCACTCAGACTAC..... 621
 : : : : :
 1101 lAlaGlnLeuAlaGlnGluIlePheArgSerHisLeuGluHisGlnAspT 1118
 : : : : :
 622ATCTGCCAGCGCCACTTCCAGGACCCAGGA 652
 : : : : :
 1118 hrLeuLeuLysProSerGluArgThrSerProValThrLeuSerPro 1134
 : : : : :
 653 CCATCAIT.....CAGAAGGAGACCCATT..... 675
 : : : : :
 1135 HisLysHisValSerGlyPheSerSerSerSerLeuArgThrSerSerThrGl 1151
 : : : : :
 676GACCTCGGGTCAAGGCCACC... 696
 : : : : :
 1151 yAspAlaGlyGlySerArgArgProHisArgLysProThrIleLeuAla 1168

```

1413 CGAACGCTTCTTCCCTAT.....GCCAATGGACCCCTGGGCATTC 1415
1457 laAsnLeuSerGlyGlySerGlnGlyGluPheSerCysLeuAlaGlnAsn 1473
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1454 GAGACCTCCAGGCCAATGACACCGGACGCTACTTCTGCTGGCTGCCAAT 1503
1474 GluAlaGlyValLeuMetGlnTysAlaSerLeuValIleGlnAspTyrTr 1490
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1504 GACCAAAACAANTGTTACCATCATGGCTACCTGAAGTTAAAGAT..... 1548
1490 pTrpSerValAspArgLeuAlaThrCysSerAlaSerCysGlyAsnArgG 1507
1549 .....GCAACTCAGA 1558
1507 lyValGlnGlnProArgLeuArgCysLeuLeuAsnSerThrGluValAsn 1523
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1559 TCACCTCAGGGGCCCCGAGCACCAATCGAGAAGAAAGGTTCCAGGGTGACC 1608
1524 ProAlaHisCysAlaGlyLysValArgProAlaValIleGlnProIleAlaCys 1540
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1609 ...TTCACGTGCCAGGCTCCTTTGACCCCTCCTTGACGCCCATCAC 1655
1540 sAsnArgArgAspCys..... 1545
1566 CTGGCGTGGGACGCTCGAGACCTCCAGAGCTTGGGGACAGTCACAAGT 1705
1546 .....ProSerArg...TrpMetValThrSerTrpSerAlaCysTh 1558
1706 ACTTCATAGAGGATGGCGCTGTGTCATCCACAGCTGGACTACAGGCAC 1755
1558 rArgSerCysGlyGlyValGlnThrArgArgValThrCysGlnLysL 1575
1756 CAGG..... 1759
1575 euLysAlaSerGlyIleSerThrProValSerAsnAspMetCysThrGln 1591
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1760 .....GCAACTACAGCTGCGGCGCCAGTACCGAACTGGATGCTGTGGAGA 1804
1592 ValAlaLysArgProValAspThrGlnAlaCysAsnGlnGlnLeuCysVa 1608
1805 GTAGGGCACAGC..... 1816
1608 lGluTrpAlaPheSerSerTrpGlyGlnCysAsnGlyProCys 1622
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1817 .TCTTGTGTGGGGAGCCCTGGCGCGTGCACAGCGCTGGTG 1858
seq_name: /cgn2_6/ptodata/2/ina/6B_COMB.seq:US-08-506-296B-13
seq_documentation_block:
; Sequence 13, Application US/08506296B
; Patent No. 6313265
; GENERAL INFORMATION:
; APPLICANT: Phillips, Greg
; APPLICANT: Cunningham, Bruce A.
; APPLICANT: Crossin, Kathryn L.
; TITLE OF INVENTION: NEURITE OUTGROWTH-PROMOTING POLYPEPTID
; NUMBER OF SEQUENCES: 77
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: The Scripps Research Institute
; STREET: 10550 No. 6313265th Torrey Pines Road, TPC-8
; CITY: La Jolla
; STATE: California
; COUNTRY: U.S.
; ZIP: 92037
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/506, 296B
; FILING DATE: 24-JUL-1995

```

```

/ CLASSIFICATION: 514
/ ATTORNEY/AGENT INFORMATION:
/ NAME: Fitting, Thomas
/ REGISTRATION NUMBER: 34,163
/ REFERENCE/DOCKET NUMBER: TSI 488.0
/ TELECOMMUNICATION INFORMATION:
/ TELEPHONE: (619) 554-2937
/ TELEFAX: (619) 554-6312
/ INFORMATION FOR SEQ ID NO: 13:
/ SEQUENCE CHARACTERISTICS:
/ LENGTH: 3888 base pairs
/ TYPE: nucleic acid
/ STRANDEDNESS: double
/ TOPOLOGY: linear
/ MOLECULE TYPE: cDNA
/ HYPOTHETICAL: NO
/ ANTI-SENSE: NO
/ FEATURE:
/ NAME/KEY: CDS
/ LOCATION: 12..3773
/ US-08-506-296B-13

```

```

alignment_scores:
  Quality: 312.00      Length: 815
  Ratio: 0.904        Gaps: 30
  Percent Similarity: 42.331  Percent Identity: 20.368

alignment_block:
US-10-044-807-2 x US-08-506-296B-13 ..
Align seg 1/1 to: US-08-506-296B-13 from: 1 to: 3888

843 ProCysMetLeuAlaThrCysAlaArgProGlyArgProSerThrLysHi 859
|||||
57 CCGTCCGCTGCTTATCCAGATC.....CCCGAGGAATATGA 91

859 sSerProHisIleAlaAlaArgLysValTyrIleGlnThrArgArgG 876
|||||
92 AGGACACCATGTGTAGGACCACTGTCATCAGGAAACAGTCTCCACGGC 141

876 nArgLysLeuHisPheValValGlyGlyPheAlaTyrLeuLeuProLys 892
|||
142 GC.....CTGGTGTCTTCCCCACA 161

893 ThrAlaValValLeuArgCysProAlaArgValArgLysProLeuIl 909
|||||
162 GATGACATCAGCTCAAGTGTGAGGCCAGTGGCAAGCCCAAGTGCAGTT 211

909 eThrTrpGluLysAspGlyGlnHisIleLeuSerSerThrHisValThrV 926
|||
212 CCGCTGGACGAGGATGTGTCTCACTTCAACCCCAAGAGAGCTGGGTG 261

926 alAlaPropheGlyTyrLeuLysIleHisArgLeuLysProSerAspAla 942
|||||
262 TGACCGGTGAC.....CAGTCCGCCCATCTCT 287

943 GlyValThrCysSerAlaGlyProAlaArgGluHisPheValIleLy 959
|||||
288 GGCCTCTTCAACATCAGGGCAAC.....AACAGCAACTTGTCTCAGAG 331

959 sLeuIleGlyGlyAsnArgLysLeuValAlaArgProLeuSerProArgS 976
|||||
332 GTTCCAGGGCATCTACCGCTGCTTTGGCCAGCAATAGCTGGGCACCGCCA 381

976 erGluGluValLeuAlaGlyArgLysGlyGlyProLysGluAlaLeu 992
|||||
382 TGTCCCATGAGATCCGGCTCATGCCGAGGCTCCCCCAAG..... 422

993 GlnThrHisLysHisGlnAsnGlyIlePheSerAsnGlySerLysLag1 1009
|||||
422 .....

```

```

1009 uLysArgGlyLeuAlaAlaAsnProGlySerArgTyrAspLeuValS 1026
422 .....
1026 erArgLeuLeuGlnGlyTyrProGlyGluLeuLeuAlaSerTirp 1042
|||||
423 .....TGGCAAAGGAGACAGTGAAGCCCGTG 449
1043 GluAlaGlnAspSerAlaGluArgAsnThrThrSerGluGluAspProgl 1059
|||||
450 GAGGTGGAGGAGGGAGTCACTGGTTCTGCCTTGCACCCCTCCCCCAAG 499
1059 yAlaGluGlnVal.....LeuLeuHisLeuP 1068
|||||
500 TGCAGAGCCTCTCCGATCTACTGGATGAACAGCAAGATCTTGCACATCA 549
1068 roPheThrMetValThrGluGlnArgArgLeuAspIleLeuGlyAsn 1084
|||||
550 AG.....CAGGACGAGCGGTGACGATGGCCAGACACGCCAAC 587
1085 LeuSerGlnGlnProGluGluLeuArgAspLeuTyrSerLysHisLeuVa 1101
|||||
588 CTCTACTTTGCCAATGTGCTCACTCCGACAACTCAGACTAC..... 632
1101 lAlaGlnLeuAlaGlnGluIlePheArgSerHisLeuGluHisGlnAspT 1118
|||||
633 .....ATGTGCCACGCCCTTCCAGGCACGAGCA 663
1118 hrLeuLeuLysProSerGluArgArgThrSerProValThrLeuSerPro 1134
|||||
664 CCATCATT.....CAGAAGGAACCCATT..... 686
1135 HisLysHisValSerGlyPheSerSerLeuArgThrSerSerThrGl 1151
|||||
687 .....GACCTCCGGGTCAAGGCCACC... 707
1151 yAspAlaGlyGlySerArgArgProHisArgLysProThrIleLeuA 1168
|||||
708 .....ACAGCATGATTGACAGGAGCCGCGCTGTCT 741
1168 rGLysIleSerAlaAlaGlnGlnLeuSerAlaSerGluValValThrHis 1184
|||||
742 TCCCCACCAAC.....TCCAGCAGCCACCTGGTGGCCTTG 776
1185 LeuGlyGlnThrValAlaLeuAlaSerGlyThrLeuSerValLeuLeuHi 1201
|||||
777 CAGGGGAG.....CCATTGGTCTCTGA 799
1201 sCysGluAlaIleGlyHisProArgProThrIleSerTirpAlaArgAsnG 1218
|||||
800 GTGCATCGCCGAGGCTTCCACGCCCCCATCAATGGTGGCGCCCA 849
1218 lyGluGluValGlnPheSerAspArgIleLeuLeuGln...ProAspAsp 1233
|||||
850 GTGCCCCCATGCCA...GCCAGCGTGTCACTACCAGAACCAACAAG 896
1234 SerLeuGlnIleLeuAlaProValGluAlaAspValGlyPheThrCy 1250
|||||
897 ACCCTGACGTCTGTGAAGTGGCGAGGAGGATGATGGCGAGTACCGCTG 946
1250 sAsnAlaThrAsnAlaLeuGlyTyrAspSerValSerIleAlaValThrL 1267
|||||
947 CCTGGCCGAGAACTCACTGGGCAGTGGCCGATGCGTACTATGTCTACCG 996
1267 euAlaGlyLysProLeuValLysThrSerArgMetThrValIleAsnThr 1283
|||||
997 TGGAGGTGCCCGTACTGG.....CTG 1019
1284 GluLysProAlaValThrValAspIleGlySerThrIleLysThrValGl 1300
|||||
1020 CACAAGCCC.....CAGAGCCATCTATATGGGCC 1048
1300 nGlyValAsnValThrIleAsnCysGlnValAlaGlyValProGluAlaG 1317

```

Wed Jul 24 11:35:44 2002

```
1049 AGGAGAGACTGCCGCTGCAGCTGCAAGTCCAGGGCAGGCCCAACACAG 1098
||||| : : : : : : : : : : : : : : : : : : : : : : : :
1317 luValThrTrpPheArgAsnLysSerLysLeuGlySerPro..... 1330
||||| : : : : : : : : : : : : : : : : : : : : : : : :
1099 AGGTCACCTGGAGAATCAAC.....GGGATCCCTGTGGAGGAG 1136
||||| : : : : : : : : : : : : : : : : : : : : : : : :
1331 .....HisHisLeuHisGluGlySerLeuLeuLeuTh 1341
||||| : : : : : : : : : : : : : : : : : : : : : : : :
1137 CTGGCCAMAGACCAAGATACCGATTCACGCTGGCCCTGATCTGAG 1186
||||| : : : : : : : : : : : : : : : : : : : : : : : :
1341 rAsnValSerSerAspGlnGlyLeuThrSerCysArgAlaAlaAsnL 1358
||||| : : : : : : : : : : : : : : : : : : : : : : : :
1187 CAACGTGACGCCAGTCACACATGGTGACCAATGTGAGGCCGCCAACC 1236
||||| : : : : : : : : : : : : : : : : : : : : : : : :
1358 euHisGlyGluLeuThrGlnLeuSerThrGlnLeuLeuLeuLeuAspPro 1374
||||| : : : : : : : : : : : : : : : : : : : : : : : :
1237 GGCAGGGCTCTTCTGGCCCAATGCCATCTACGTGTGC..... 1277
||||| : : : : : : : : : : : : : : : : : : : : : : : :
1375 GlnValProThrGlnLeuGluAspIleArgAlaLeuLeuAlaThrGl 1391
||||| : : : : : : : : : : : : : : : : : : : : : : : :
1278 CAGTGCACGCCAG..... 1292
1391 yProAsnLeuProSerValLeuThrSerProLeuGlyThrGlnLeuValL 1408
||||| : : : : : : : : : : : : : : : : : : : : : : : :
1293 .....ATCCTGACTGGGCAATCAGACGTACATGCGTG 1326
||||| : : : : : : : : : : : : : : : : : : : : : : : :
1408 euAspProGlyAsnSerAlaLeuLeuGlyCysProIleLysGlyHisPro 1424
||||| : : : : : : : : : : : : : : : : : : : : : : : :
1327 TCCAG...GGCAGCAGCTGCCATCTCTGTCGAAGCCCTCGGAGCGCT 1373
||||| : : : : : : : : : : : : : : : : : : : : : : : :
1425 ValProAsnIleThrTrpPhe...HisGlyGlnProIleValThrAl 1440
||||| : : : : : : : : : : : : : : : : : : : : : : : :
1374 GTGCCAGTGTTCAGTGGCTGGACGAGGTGGCAACAGTGTTCAGGA 1423
||||| : : : : : : : : : : : : : : : : : : : : : : : :
1440 aThrGlyLeuThrHisHisIleLeuAlaGlyGlnIleLeuGlnVala 1457
||||| : : : : : : : : : : : : : : : : : : : : : : : :
1424 CGAACGCTTCTTCCCTAT.....GCCAATGGGACCTGGGCATTC 1464
||||| : : : : : : : : : : : : : : : : : : : : : : : :
1457 laAsnLeuSerGlyCysSerGlnGlyGluPheSerCysLeuAlaGlnAsn 1473
||||| : : : : : : : : : : : : : : : : : : : : : : : :
1465 GAGACCTCCAGGCCCAATGACACCGGACCTACTTCTGCTGGCTGCCAAT 1514
||||| : : : : : : : : : : : : : : : : : : : : : : : :
1474 GluAlaGlyValLeuMetGlnLysAlaSerLeuValIleGlnAspTrp 1490
||||| : : : : : : : : : : : : : : : : : : : : : : : :
1515 GACCAAAACATGTTACCATCATGCTTAACCTGAAGGTTAAAGAT..... 1559
||||| : : : : : : : : : : : : : : : : : : : : : : : :
1490 pTrpSerValAspArgLeuAlaThrCysSerAlaSerCysGlyAsnArg 1507
||||| : : : : : : : : : : : : : : : : : : : : : : : :
1560 .....GCAACTCAGA 1569
1507 lyValGlnGlnProArgLeuArgCysLeuLeuAsnSerThrGluValAsn 1523
||||| : : : : : : : : : : : : : : : : : : : : : : : :
1570 TCACTCAGGGCCCGCCAGCACAATGACAGCAAGAGTTCCAGGTGACC 1619
||||| : : : : : : : : : : : : : : : : : : : : : : : :
1524 ProAlaHisCysAlaGlyValArgProAlaValGlnProIleAlaCy 1540
||||| : : : : : : : : : : : : : : : : : : : : : : : :
1620 ....TTCAGTGGCAGGCTCCTTTCACCCCTCTTGCAGCCCGCAGCATCAC 1666
||||| : : : : : : : : : : : : : : : : : : : : : : : :
1540 sAsnArgArgAspCys..... 1545
||||| : : : : : : : : : : : : : : : : : : : : : : : :
1667 CTGGGTGGGACGCTCGAGACCTCCAGGAGCTTGGGACAGTGACAAGT 1716
||||| : : : : : : : : : : : : : : : : : : : : : : : :
1546 .....ProSerArg.TrpMetValThrSerTrpSerAlaCysTh 1558
||||| : : : : : : : : : : : : : : : : : : : : : : : :
1717 ACTTCATAGAGGATGGCCCGCTGGTCATCCACAGCTGGACATACAGCGAC 1766
||||| : : : : : : : : : : : : : : : : : : : : : : : :
1558 rArgSerCysGlyGlyValGlnThrArgValThrCysGlnLysL 1575
||||| : : : : : : : : : : : : : : : : : : : : : : : :
1767 CAGG..... 1770
||||| : : : : : : : : : : : : : : : : : : : : : : : :
1575 euLysAlaSerGlyIleSerThrProValSerAsnAspMetCysThrGln 1591
||||| : : : : : : : : : : : : : : : : : : : : : : : :
```

```

391 .....ACAGCAAAAGTTGAGTAGCAGGACCACCTGAGGTTCCCTTCAC 433
1178 laSerGluValValThr...HisLeuGlyGlnThrValAlaLeuAlaSer 1193
434 ACAGCAATCTCTCACAGCCTTCATGGGAGACACA..... 468
1194 GlyThrLeuSerValLeuLeuHisCysGluAlaLeuGlyHisProArgPr 1210
469 .....GTGCTACTCAAGTGTGAAGTCATTGGGAGCCCATGCC 506
1210 oThrIleSerTrpAlaArgAsnGlyGlu.....ValGlnPheS 1224
507 AACAACTCACTGGCAGAGAACCAACAGACCTGACTCCATCCAGGTG 556
1224 erAspArgIleLeuLeuGlnProAspAspSerLeuGlnIleLeuAlaPro 1240
557 ACTCCGAGTGGTGGCTTCGCCCTCTGGAGCATTCGACATTCGACGCGACTC 606
1241 ValGluAlaAspValGlyPheThrCysAsnAlaThrAsnAlaLeuG 1257
607 CAACCGGGGACATTTGAATTTACCGATGTCAGCTCGAAATCCA..... 651
1257 yTrpSerValSerIleAlaValThrLeuAlaGlyLysProLeuValL 1274
652 .....GCCAGCTCAAGACAGGA..... 669
1274 yThrSerArgMetThrValIleAsnThrGluLysProAlaValThrVal 1290
670 .....AATGAAGCAGAAGTCAGAATTTTATCA 696
1291 AspileGly.....SerThrII 1296
697 GATCCAGGACTGCATACACACCTGTATTTCGCAAGACCATCCAATGT 746
1296 eLysThrValGlnGlyValAsnValThrIleAsnCysGlnValAlaGlyV 1313
747 AGTAGCCATTGAAGGAAGATGCTGCCGGAATGTGTGTTCGGCT 796
1313 alProGluAlaGluValThrTrpPheArgAsnLysSerLysLeu..... 1327
797 ATCCTCCACCAAGATTACTGTGTTAGGCGGAGGAAGTCATCCAATC 846
1328 ...GlySerProHisHisLeuHisGluGlySer...LeuLeuThrAs 1342
847 AGGTCTAAAAAGTATCTTTATTGGGTGAAGCAACTTGTATTCTCCAA 896
1342 nValSerSerAspGlnGlyLeuTyrSerCysArgAlaAlaAsnLeuH 1359
897 TGTGACAGATGATGACAGTGAATGTATACCTGTGTGTGCACATATAAAA 946
1359 isGlyGluLeuThrGluSerThrGlnLeuLeuIleLeuAspProGln 1375
947 ATCAGATATATTAGTCCTCTGCAGAGCTCACAGTCTGTGGTCCGCCATGG 996
1376 ValProThrGlnLeuGluAspIleArgAlaLeuLeuAlaAlaThrGlyPr 1392
997 TTT.....TT 1001
1392 oAsnLeuProSerValLeuThrSerProLeuGlyThrGlnLeuValLeuA 1409
1002 AAATCATCTCTCAACCTGTATGCTATGAAGCATGGATATTGAGTTTG 1051
1409 spProGlyAsnSerAlaLeuLeuGlyCysProIleLysGlyHisProVal 1425
1052 AA.....TGTACAGCTCTCTGGAAGCCTGTG 1077
1426 ProAsnIleThrTrpPheHisGlyGlnProIleValThrAlaThrG 1442
1078 CCCACTGTCAATTGGATGAAGATGGAGATGGTCAATCTTAGT..... 1122
1442 yLeuThrHisHisIleLeuAlaAlaGlyGlnIleLeuGlnValAlaAsnL 1459
1123 ....GATTATTTTCAGATAGTGGGAGGAACCACTTACGGATACTTGGG 1168

```

```

1459 euSerGlyGlySerGlnGlyGluPheSerCysLeuAlaGlnAsnGluAla 1475
1169 TGGTGAAGTCAGATGAAGCCTTTTATCAATGTGTGGCTGAAATGAGGCT 1218
1476 GlyValLeuMetGlnLysAlaSerLeuValIleGlnAspTyrTrpTrpSe 1492
1219 GGAATGCCAGACCCAGCTGCACAGCTCATTTGTCCT..... 1254
1492 rValAspArgLeuAlaThrCysSerAlaSerCysGlyAsnArgGlyValG 1509
1254 ..... 1254
1509 lnGlnProArgLeuArgCysLeuLeuAsnSerThrGluValAsnProAla 1525
1255 .....AAGCCTGCA 1263
1526 HisCysAlaGlyLysValArgProAla.....ValGlnProII 1538
1264 ATCCCAAGCTCCAGTGTCTCTCCCTCGGCTCCAGAGATGTGGTCCCTGT 1313
1538 eAlaCysAsnArgArgAspCysProSerArgTrpMetValThrSerTrpS 1555
1314 CTTGGTITTC.....AGCCGATTGTGCTCGTCTCAGCTGGC 1348
1555 erAlaCysThrArgSerCysGlyGlyGlyValGlnThr..... 1567
1349 GCCCA...CCTGCAGAAAGCGAAAGGAACATTCAAACTTTCACGCTCTTT 1395
1567 ..... 1567
1396 TTCTCCAGAGAAGGTGACACAGGAGGAGCATTTGAATACAACACAGCC 1445
1568 .....ArgArgValThrCysGlnLysLeuLysAlaSerGlyIleSerT 1582
1446 TGGTCCCTTTCAGCTCAGTGTGGGAAACCTGAGCCAGAACCCATGTACA 1495
1582 hr.....ProValSerAsnAspMet 1588
1496 CTTTCAGATTGGCTTACAATGAATGGGACCGGAGAGAGTCTCAA 1545
1589 CysThrGlnValAlaLysArgPro.....Va 1597
1546 CCCATCAAGGTGGCCACACAGCTGAGTTGCAAGTTCAGGGCCAGTAGA 1595
1597 lAspThrGlnAlaCysAsn.....GlnGlnLeuCysValGlnTrpA 1611
1596 AAACCTGCAAGCTGTATCTACCTCACCTCACTCAATTTCTATTACCTGGG 1645
1611 laPheSerSerTrpGlyGlnCysAsnGlyProCysIleGlyProHisLeu 1627
1646 AACCCCTGCTGCTAT.....GCAACGGTCCA..... 1671
1628 AlaValGlnHisArgGlnValPheCysGln..... 1637
1672 ...GTCCAAGTTACAGATTGTTCTGCAGCTGAGGTGTCCACAGAAAAGA 1718
1637 ..... 1637
1719 ACAGATATAGAGTTGATGGACTATCTTATAAATCGGAAGGCTGAAAA 1768
1637 ..... 1637
1769 AATTCACCAATATAGTCTTCGATTTAGTCTTATAATCGTATGTCGG 1818
1638 .....ThrArgAspGlyIleThr.....LeuProSe 1646
1819 GGCGTCTCTACTGATATATACAGTGGTACACTTCTGACGTGCAAG 1868
1646 rGlu.....G 1648
1869 TGCCCCGCCCTCAGAACGTCTCCCTGGGAGTGGTCAATTCAAGNAGTATCA 1918

```

Wed Jul 24 11:35:44 2002

```
1648 IncysSerAlaLeuProArgProValSerThrGlnAsnCysTrpSerGlu 1664
      :: ||| ||||| ||| ::||| ||||| ::|
1919 ASGTTAGTGGTGGCTCTCTCCATCAGGAACACAAAATGATTTATACC 1968
      :: ||| ||||| ||| ::||| ||||| ::|
1665 AlaCysSerValHisTrpArgVal..... 1672
      :: ::|||::| |||
1969 GGCTATAAAATTCACACACAGAGAGCCCGCAGGGTTCAGATGGAAC 2018
1673 .....SerLeuTrpThrLeuCysThrAla.....ThrCysG 1683
      ::|||::| ||| |||||
2019 ACTGGAGCAACAACTCTGGTACCTATTCACAGGACTGGGAAAGGAA 2068
      ::|||::| ||| |||||
1683 lYAsnTrpGlyPheGlnSerArgValGluCysValHisAlaArgThr 1699
      ::|||::| ||| |||||
2069 GTCAATAGTTCACAG.....GTGTGAGCCATGACA 2100
      ::|||::| ||| |||||
1700 AsnLysAlaValProGluHisLeuCysSerTrpGlyProArgProAlaAs 1716
      ::|||::| ||| |||||
2101 GTCAATAGTTCACAG.....GGACCA...CCTTCCAA 2126
      ::|||::| ||| |||||
1716 nTrpGlnArgCysAsnIleThrProCysGluAsnMetGlu 1729
      ::|||::| ||| |||||
2127 CTGTTACTGACAG...ACTCAGAGATGATCTAGAT 2163
```

seq_name: /cgn2_6/ptodata/2/ina/PCTUS_COMB.seq:PCT-US94-05277-1

```
seq_documentation_block:
; Sequence 1, Application PC/TUS9405277
; GENERAL INFORMATION:
; APPLICANT: Bruskin, Arthur
; APPLICANT: Jarosz, David E.
; APPLICANT: Johnson, Karen
; APPLICANT: Kinzler, Kenneth W.
; APPLICANT: Vogelstein, Bert
; APPLICANT: Zdzienicka, James R.
; TITLE OF INVENTION: Antibodies Specific for DCC Gene Product
; NUMBER OF SEQUENCES: 2
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Banner, Birch, McKie & Beckett
; STREET: 1001 G Street, N.W.
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20001
; COMPUTER READABLE FORM:
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US94/05277
; FILING DATE:
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Kagan, Sarah A.
; REGISTRATION NUMBER: 32,141
; REFERENCE/DOCKET NUMBER: 01107.42709
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202.508.9100
; TELEFAX: 202.508.9299
; TELEX: 197430 BMB UT
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 4608 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: Homo sapiens
; POSITION IN GENOME:
```

```
; CHROMOSOME/SEGMENT: 18q21
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..4342
; PCT-US94-05277-1

alignment_scores:
  Quality: 311.50      Length: 730
  Ratio: 0.986        Gaps: 32
  Percent similarity: 43.288      Percent Identity: 20.685

alignment_block:
US-10-044-807-2 x PCT-US94-05277-1 ..
Align seg 1/1 to: PCT-US94-05277-1 from: 1 to: 4608

1148 SerSerThrGlyAspAlaGlyGlySerArgArgProHisArgLysPr 1164
      ::|||::| ||| |||||
355 GCATCTTTAGGAGATTCTGGCTCAATTATTAGTCGG..... 390
      ::|||::| ||| |||||
1164 oThrIleLeuArgLysIleSerAlaAla.....GlnGlnLeuSerA 1178
      ::|||::| ||| |||||
391 .....ACAGCAAAAGTTGCAGTAGCAGGACCACCTGAGGTTCTTTCAC 433
      ::|||::| ||| |||||
1178 laSerGluValValThr...HisLeuGlyGlnThrValAlaLeuAlaser 1193
      ::|||::| ||| |||||
434 AGACAGAATCTGTACAGCCCTTCATGGGAGACACA..... 468
      ::|||::| ||| |||||
1194 GlyThrLeuSerValLeuHisCysGluAlaIleGlyHisProArgpr 1210
      ::|||::| ||| |||||
469 .....GTGCTACTCAAGTGTGAAGTCATTTGGGAGCCCATGCC 506
      ::|||::| ||| |||||
1210 oThrIleSerTrpAlaArgAsnGlyGlu.....ValGlnPheS 1224
      ::|||::| ||| |||||
507 AACATCCACTGGCAGAAAGAACCAAGACCTGACTCCATCCGAATCCA..... 551
      ::|||::| ||| |||||
1224 erAspArgIleLeuLeuGlnProAspSerSerLeuGlnIleLeuAlaPro 1240
      ::|||::| ||| |||||
557 ACTCCGAGTGTGGTCTTGCCTCTGGAGCATTCAGATCAGCCGAGCTC 606
      ::|||::| ||| |||||
1241 ValGluAlaAspValGlyPheTrpCysAsnAlaThrAsnAlaLeuGl 1257
      ::|||::| ||| |||||
607 CAACCGGGGACATTTGGAATTTACCGATGCTCAGCTCGAATCCA..... 651
      ::|||::| ||| |||||
1257 yTrAspSerValSerIleAlaValThrLeuAlaGlyLysProLeuValL 1274
      ::|||::| ||| |||||
652 .....GCCAGCTCAAGAACAGGA..... 669
      ::|||::| ||| |||||
1274 ysThrSerArgMetThrValIleAsnThrGluLysProAlaValThrVal 1290
      ::|||::| ||| |||||
670 .....AATGAGCAGAAGTCAGAATTTATCA 696
      ::|||::| ||| |||||
1291 AspIleGly.....SerThrI 1296
      ::|||::| ||| |||||
697 GATCCAGGACTCCATAGACAGCTGATTTTTCGAAAGACCACCAATGT 746
      ::|||::| ||| |||||
1296 eLysThrValGlnGlyValAsnValThrIleAsnCysGlnValAlaGlyV 1313
      ::|||::| ||| |||||
747 AGTAGCCATTGAAGGAAAAGATGCTGCTCGAATGTTCTGTCTGCT 796
      ::|||::| ||| |||||
1313 alProGluAlaGluValThrTrpPheArgAsnLysSerLysLeu..... 1327
      ::|||::| ||| |||||
797 ATCTCCACCACCAAGTTTACCTGGTTACGAGGGGAGGAGTCATCCAACTC 846
      ::|||::| ||| |||||
1328 ...GlySerProHisHisLeuHisGluGlySer...LeuLeuLeuThrAs 1342
      ::|||::| ||| |||||
847 AGGTCTFAAAAGATATCTTTTATTGGGTGGAAGCAACTTGCTTATCTCAA 896
      ::|||::| ||| |||||
1342 nValSerSerSerAspGlnGlyLeuTyTrSerCysArgAlaAlaAsnLeuH 1359
      ::|||::| ||| |||||
897 TGTGACAGATGATGACAGTGAATGTATACCTGTGTGTGCACATATAAAA 946
      ::|||::| ||| |||||
```


Wed Jul 24 11:35:44 2002

```

; FILING DATE: 24-JUL-1995
; CLASSIFICATION: 514
; ATTORNEY/AGENT INFORMATION:
; NAME: Fitting, Thomas
; REGISTRATION NUMBER: 34,163
; REFERENCE/DOCKET NUMBER: TSRI 488.0
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (619) 554-2937
; TELEFAX: (619) 554-6312
; INFORMATION FOR SEQ ID NO: 27:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3943 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; HYPOTHEetical: NO
; ANTI-SENSE: NO
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 33..3839
; US-08-506-296B-27

```

```

alignment_scores:
  Quality: 305.00      Length: 955
  Ratio: 0.723         Gaps: 41
  Percent Similarity: 44.188      Percent Identity: 20.105

```

alignment_block:

US-10-044-807-2 x US-08-506-296B-27 ..

Align seg 1/1 to: US-08-506-296B-27 from: 1 to: 3943

```

891 ProLysThrAlaValValLeuArgCysProAlaArgValArgLysPr 907
||||| :|||:|||||:|||||:|||||:|||||:|||||:|||||:
195 CCTCAGAGAAATATGTAATACATATGTAAGCAAAAGGAAACCCCTCC 244
||||| :|||:|||||:|||||:|||||:|||||:|||||:
907 oLeuileThrTrpGluLysAspGlyGlnHis.....LeuileSerSert 922
| :|||:|||||:|||||:|||||:|||||:|||||:
245 TAGCTTCTCTGGACGGCGCAATGGAATCATTTTGATATAGATAAAGATG 294
||||| :|||:|||||:|||||:|||||:|||||:|||||:
922 hrHisValThrValAlaPro...PheGlyTyrLeuLysIleHisArgLeu 937
:|||||:|||||:|||||:|||||:|||||:|||||:
295 CACAGGTAAACAATGAAACCAATTCAGGAACCCCTGTTGTAATATATATG 344
||||| :|||:|||||:|||||:|||||:|||||:|||||:
938 LysProSerAspAla.....GlyValTyrThrCysSerAlaG1 950
:|||||:|||||:|||||:|||||:|||||:|||||:
345 AATGGTGTGAAGGAGAGCATATGAAAGGATATACCAAGTGTACAGCAG 394
||||| :|||:|||||:|||||:|||||:|||||:|||||:
950 yProAlaArgGluHisPheValIleLysLeuileGlyGlyAsnArgLysL 967
||||| :|||:|||||:|||||:|||||:|||||:|||||:
395 GAATCAAGAGGAGCAGCCATT.....TCCAACAATA 426
||||| :|||:|||||:|||||:|||||:|||||:|||||:
967 euValAlaArgPro...LeuSerProArgSerGluGluValLeuAla 982
:|||||:|||||:|||||:|||||:|||||:|||||:
427 TTGTTATACGGCCATCTAGATCCCGTTG..... 455
||||| :|||:|||||:|||||:|||||:|||||:|||||:
983 GlyArgLysGlyGlyProLysGluAlaLeuGlnThrHisLysHisGlnAs 999
||||| :|||:|||||:|||||:|||||:|||||:|||||:
456 .....TGGACTAAAGAAATACTAGAACCAATCATGTTCCGAGA 493
||||| :|||:|||||:|||||:|||||:|||||:|||||:
999 nGlyIlePheSerAsnGlySerLysAlaGluLysArgGlyLeuAlaAla 1016
||||| :|||:|||||:|||||:|||||:|||||:|||||:
494 AGGT..... 497
||||| :|||:|||||:|||||:|||||:|||||:|||||:
1016 snProGlySerArgTyrAspAspLeuValSerArgLeuLeuGluGlnGly 1032
||||| :|||:|||||:|||||:|||||:|||||:|||||:
498 .....GATTCCCTTAGTACTAACTGCAGACCTCCTGTT 530
||||| :|||:|||||:|||||:|||||:|||||:|||||:
1033 GlyTrpProGlyGluLeuLeuAlaSerTrpGluAlaGlnAspSerAlaG1 1049
||||| :|||:|||||:|||||:|||||:|||||:|||||:
531 GGCTTACCACCACTATAATA...TTTGG.....ATGGATAATGCTTT 571
||||| :|||:|||||:|||||:|||||:|||||:|||||:

```

```

1049 uArgAsnThrThrSerGluGluAspProGlyAlaGluGlnValLeuLeuH 1066
||||| :|||:|||||:|||||:|||||:|||||:|||||:
572 CCAAGGCGTG.....CCTCAAGTGAAAGAGTT..... 599
||||| :|||:|||||:|||||:|||||:|||||:|||||:
1066 isLeuProPheThrMetValThrGluGlnArgArgLeuAspAspIleLeu 1082
||||| :|||:|||||:|||||:|||||:|||||:|||||:
600 .....TCTCAAGGTCTCAATGGAGACCTT 623
||||| :|||:|||||:|||||:|||||:|||||:|||||:
1083 GlyAsnLeuSerGlnProGluGluLeuArgAspLeuTyrSerLysHi 1099
||||| :|||:|||||:|||||:|||||:|||||:|||||:
624 TATTTTCTTAATGTACAAACAGAGGACACCGTGTGGACTAT..... 665
||||| :|||:|||||:|||||:|||||:|||||:|||||:
1099 sLeuValAlaGlnLeuAlaGlnGluIlePheArgSerHisLeuGluHisG 1116
||||| :|||:|||||:|||||:|||||:|||||:|||||:
666 .....ATCTGCTACGCGAGATTTAATCACA 690
||||| :|||:|||||:|||||:|||||:|||||:|||||:
1116 lnAspThrLeuLysProSerGluArgArgThrSerProValThrLeu 1132
||||| :|||:|||||:|||||:|||||:|||||:|||||:
691 CACAAACTATA.....CAGCAGAAACAACCCATTCT... 722
||||| :|||:|||||:|||||:|||||:|||||:|||||:
1133 SerProHisLysHisValSerGlyPheSerSerSerLeuArgThrSerSe 1149
||||| :|||:|||||:|||||:|||||:|||||:|||||:
723 .....GTAAAGTCTTTTCAACCAAGCAGCTTACA..... 752
||||| :|||:|||||:|||||:|||||:|||||:|||||:
1149 rThrGlyAspAlaGlyGlySerArgProHisArgLysProThrI 1166
||||| :|||:|||||:|||||:|||||:|||||:|||||:
753 .....GAAAGGCCACCAAGTTC 768
||||| :|||:|||||:|||||:|||||:|||||:|||||:
1166 leLeuArgLysIleSerAlaAlaGlnGlnLeuSerAlaSerGluValVal 1182
||||| :|||:|||||:|||||:|||||:|||||:|||||:
769 TTCIT..... 773
||||| :|||:|||||:|||||:|||||:|||||:|||||:
1183 ThrHisLeuGlyGlnThr.....ValAlaLeuAlaSerGlyThrLe 1196
||||| :|||:|||||:|||||:|||||:|||||:|||||:
774 ACACCAATGGCGCAGCACAAAGTAACAAGTGGAACTGAGAGGAAATGTTCT 823
||||| :|||:|||||:|||||:|||||:|||||:|||||:
1196 uSerValLeuLeuHisCysGluAlaIleGlyHisProArgProThrIles 1213
||||| :|||:|||||:|||||:|||||:|||||:|||||:
824 T.....TTGTTGAATGATCGCAGCAGGATTCACCAACACCAAGTAAIC 867
||||| :|||:|||||:|||||:|||||:|||||:|||||:
1213 erTrpAlaArgAsnGlyGluValGlnPheSerAspArgIleLeuLeu 1229
||||| :|||:|||||:|||||:|||||:|||||:|||||:
868 GCTGGATTAAAGAGGGTGGTGAACCTGCCACCAACAGAACGTTTGTGAA 917
||||| :|||:|||||:|||||:|||||:|||||:|||||:
1230 GlnProAspAspSerLeuGlnIleLeuAlaProValGluAlaAspValG1 1246
||||| :|||:|||||:|||||:|||||:|||||:|||||:
918 AATTTTAAAGAAACTCTCAAGATTATAGACGCTCTCTGAAGCTGACTCTGG 967
||||| :|||:|||||:|||||:|||||:|||||:|||||:
1246 yPheTyrThrCysAsnAlaThrAsnAlaLeuGlyTyrAspSerValSerI 1263
||||| :|||:|||||:|||||:|||||:|||||:|||||:
968 GAACATACAAATGTACAGCAAGAAATACATITGGTTCTACTCATCATGCA 1017
||||| :|||:|||||:|||||:|||||:|||||:|||||:
1263 leAlaValThrLeuAlaGlyLysProLeuValLysThrSerArgMetThr 1279
||||| :|||:|||||:|||||:|||||:|||||:|||||:
1018 TTTCGGTAACTGTAAGCTGCCCATATACTGGATTAACAGACCCAGGAAC 1067
||||| :|||:|||||:|||||:|||||:|||||:|||||:
1280 ValIleAsnThrGluLysProAlaValThrValAspIleGlySerThrI1 1296
||||| :|||:|||||:|||||:|||||:|||||:|||||:
1068 TTAGTATTGTCT..... 1079
||||| :|||:|||||:|||||:|||||:|||||:|||||:
1296 elYstThrValGlnGlyValAsnValThrIleAsnCysGlnValAlaGlyV 1313
||||| :|||:|||||:|||||:|||||:|||||:|||||:
1080 .....CCTGGAGAAGATGGGACATTGATCTCCAGAGCTAATGGCA 1119
||||| :|||:|||||:|||||:|||||:|||||:|||||:
1313 alProGluAlaGluValThrTrpPheArgAsnLysSerLysLeuGly... 1328
||||| :|||:|||||:|||||:|||||:|||||:|||||:
1120 ACCCAAAACCTAGCATAGCTGTTTAAACAATGGCGTTCCTCATAGCAATT 1169
||||| :|||:|||||:|||||:|||||:|||||:|||||:
1329 .....SerProHisLysLeuHisGluGly...SerLeuLeuLeuH 1341
||||| :|||:|||||:|||||:|||||:|||||:|||||:
1170 GCCCCAGAAAGATCTAGCAGAAAGGTAGATGGGGATACCATTAATTTCTC 1219
||||| :|||:|||||:|||||:|||||:|||||:|||||:

```

```
1341 rAsnValSerSerAspGlnGlyLeuThrSerCysArgAlaAlaAsnL 1358
: |||:||||:||||:||||:||||:||||:||||:||||:||||:
1220 AGCTGTGAAGAACGGTCAAGTGCCTGTTTATCAGTCAATGCTTCTAATG 1269
1358 euHisGlyGluLeuThrGluSerThrGlnLeuLeuLeuLeu...AspPro 1373
: |||:||||:||||:||||:||||:||||:||||:||||:||||:
1270 AGTATGGATACCTGCTGGCAATGATTTGGAATGTTCTTCTGCTGAGCCA 1319
1374 ProGlnValProThrGlnLeuLeuLeuLeuLeuLeuLeuLeuAlaAla 1390
||||:||||:||||:||||:||||:||||:||||:||||:||||:
1320 CCAAGGATCTTAACCTCCTGCTAATAAATCAATCAAGTCAATGCCA.... 1364
1390 rGlyProAsnLeuProSerValLeuThrSerProLeuGlyThrGlnLeu 1407
||||:||||:||||:||||:||||:||||:||||:||||:||||:
1365 .....GATAGTCT ..... 1373
1407 alLeuAspProGlyAsnSerAlaLeuLeuGlyCysProIleLysGlyHis 1423
||||:||||:||||:||||:||||:||||:||||:||||:||||:
1374 .....GCATTAAATAGACTGTCATTATTGTTGTTCA 1403
1424 ProValProAsnIleThrTrpPheHisGlyGlnProIleValThrAl 1440
||||:||||:||||:||||:||||:||||:||||:||||:||||:
1404 CTAAGCCTGAATCGAATGTTTAGGGAGTGAAGAGTAGCATCTTGC 1453
1440 aThrGlyLeuThrHisIleLeuAlaAlaGlyGlnIleLeuGlnValA 1457
: |||:||||:||||:||||:||||:||||:||||:||||:||||:
1454 AGGA.....AATGAATATGTTTCCATGATATGGAACCTTGGAAATTC 1497
1457 laAsnLeuSerGlyGlySerGlnGlyGluPheSerCysLeuAlaGlnAsn 1473
: |||:||||:||||:||||:||||:||||:||||:||||:||||:
1498 CAGTGGCTCAGAAGGATAGTACTGCACATACACATGTTGTCAGGAAT 1547
1474 GluAlaGlyValLeuMetGlnLysAlaSerLeuValIleGlnAspTyrTr 1490
: |||:||||:||||:||||:||||:||||:||||:||||:||||:
1548 AATTAGGAGAGCGCAATGAAGTACAACCTGGAAGTTAAGAC..... 1592
1490 pTrpSerValAspArgLeuAlaThrCysSerAlaSerCysGlyAsnArg 1507
: |||:||||:||||:||||:||||:||||:||||:||||:||||:
1593 .....CCAACGATGA 1602
1507 lyValGlnGlnProArgLeuArgCysLeuLeuAsnSerThrGlnValAsn 1523
: |||:||||:||||:||||:||||:||||:||||:||||:||||:
1603 TAAATTAACAGCCACAGTACAAAGTATGATCAGAGATCTGCCAGGCTTCA 1652
1524 ProAlaHisCysAlaGlyLysValArgProAlaValGlnProIleAlaCy 1540
: |||:||||:||||:||||:||||:||||:||||:||||:||||:
1653 ...TTGAGTGTGTATATAAACATGATCTTACCTTAATACCAACAGTTAT 1699
1540 s.....AsnArgArgAspCysProSerArgTrpMetVal. 1551
: |||:||||:||||:||||:||||:||||:||||:||||:||||:
1700 ATGCTGAAAGACAATAATGAACCTACCAGATGATGAAGGTTTCTAGTTG 1749
1552 .....ThrSerTrpSerAlaCysThrArgSerCysGlyGly 1563
: |||:||||:||||:||||:||||:||||:||||:||||:||||:
1750 GTAAGACAACTTGCACATATATGAATGTAAGTATGAAGATGATGGAACA 1799
1564 GlyValGlnThrArgArgValThrCysGlnLysLeuLysAlaSerGlyII 1580
: |||:||||:||||:||||:||||:||||:||||:||||:||||:
1800 TATACTTGATAGTTATTAATACTCTGGACAGTGTTCAGCAAGTGTGT 1849
1580 eSerThrProValSerAsnAspMetCysThrGlnVal.....A 1593
: |||:||||:||||:||||:||||:||||:||||:||||:||||:
1850 GCTTACTGTTGTGCTGCTCCCACTCCAGTATCATTTAGCTGCTGGC 1899
1593 laLysArgProValAspThrGlnAlaCysAsnGln...GlnLeuCysVal 1608
: |||:||||:||||:||||:||||:||||:||||:||||:||||:
1900 CAAATCCACCGCTTGAATGGAATTCAGAGTTCAGCTAGAAAGAACATTT 1949
1609 GluTrpAlaPheSerSerTrpGlyGlnCysAsnGlyProCysIle.... 1623
: |||:||||:||||:||||:||||:||||:||||:||||:||||:
1950 GAACCTCATGGTACCAGGAGAAATAACAGTCCATCACTTCACTTCACTT 1999
1623 ..... 1623
```

```
2000 TGTGATTGACTATGAAGATGGACTACATGAGCCAGGGGTATGGCAFTACC 2049
1624 .....GlyProHisLeuAlaValGlnHisArg..... 1632
: |||:||||:||||:||||:||||:||||:||||:||||:||||:
2050 AGACGGAAGTTCCTGGATCTCATCAACAATGTCACAGTTGAAGTTGTCTCCG 2099
1633 .....GlnValPheCysGlnThrArgAspGlyIleth 1643
: |||:||||:||||:||||:||||:||||:||||:||||:||||:
2100 TATGTCNACTACTCATTCCTGGTGTGATTTGCTGCAATGAAATGGTAGAAG 2149
1643 rLeuProSerGluGlnCys.....SerAlaLeuPro. 1653
: |||:||||:||||:||||:||||:||||:||||:||||:||||:
2150 TCAGCAAGTGAACCATCTGAACAGTACCTGACAAAGTCCGCAACCCCG 2199
1654 .....ArgProValSerThrGlnAsnCysTrpSerGlu..... 1664
: |||:||||:||||:||||:||||:||||:||||:||||:||||:
2200 ATGAATAATCCTTCTAATATGTAAGGATAGGCTCGAACCCTGATAATTG 2249
1664 ..... 1664
2250 GTAATAAGTGGAGTCTTTAAAGCTTTCAGTCTAATGACAGGACT 2299
1665 .AlaCysSerValHisTrpArgVal.....SerLeuTrpThr. 1676
: |||:||||:||||:||||:||||:||||:||||:||||:||||:
2300 CCAATATAAGTCACTGCGCGCAGAGATGTTGATGATGATGACGT 2349
1677 ..LeuCysThrAla..ThrCysGlyAsn..... 1684
: |||:||||:||||:||||:||||:||||:||||:||||:||||:
2350 CCGTTTGTAGTTGCAAAACGTGCTAAATATATTGCTGCTGGTACACCACT 2399
1685 .....TyrGlyPhe..... 1687
: |||:||||:||||:||||:||||:||||:||||:||||:||||:
2400 TTTGTTCCCTATGAATAAAGTACAGGCTTTAAATGACCTGGGATATGC 2449
1688 .....GlnSerArgArgValGluCysValHisAlaArgThrAsnLysAla 1702
: |||:||||:||||:||||:||||:||||:||||:||||:||||:
2450 ACCAGGCATCAGAGGTTATTGGACATTCAGGGAAGACTTGCCATGG 2499
1703 ValProGluHisLeuCysSerTrpGlyProArgProAlaAsnTrpGlnAr 1719
: |||:||||:||||:||||:||||:||||:||||:||||:||||:
2500 TTGCTCCAGCAATGTGCGGTTTCATGTCATTAACAGCACATTTGCCAAG 2549
1719 qCysAsnIleThr 1723
: |||:||||:||||:||||:||||:||||:||||:||||:||||:
2550 GTGCACTGGGACC 2562
seq_name: /cgn2_6/ptodata/2/ina/6B_COMB.seq:US-09-383-630-1
seq_documentation_block:
; Sequence 1, Application US/09383630A
; Patent No. 6265632
; GENERAL INFORMATION:
; APPLICANT: Avner Yayon et al.
; TITLE OF INVENTION: ANIMAL MODEL FOR FIBROBLAST GROWTH
; FACTOR RECEPTOR ASSOCIATED
; CHONDRODYSPLASIA
; NUMBER OF SEQUENCES: 18
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Mark M. Friedman c/o Anthony Castorina
; STREET: 2001 Jefferson Davis Highway, Suite 207
; CITY: Arlington
; STATE: Virginia
; COUNTRY: United States of America
; ZIP: 22202
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 1.44 megabyte, 3.5" microdisk
; COMPUTER: Twinhead* slimnote-890TX
; OPERATING SYSTEM: MS DOS version 6.2,
; Windows version 3.11
; SOFTWARE: Word for Windows version 2.0 converted
; to an ASCII file
; CURRENT APPLICATION DATA:
```

```

; APPLICATION NUMBER: US/09/383,630A
; FILING DATE: 26-Aug-1999
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: <Unknown>
; FILING DATE: <Unknown>
; ATTORNEY/AGENT INFORMATION:
; NAME: Friedmann, Mark M.
; REGISTRATION NUMBER: 33,883
; REFERENCE/DOCKET NUMBER: 1402/2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 972-3-5625553
; TELEFAX: 972-3-5625554
; TELEX: <Unknown>
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 5993
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; SEQUENCE DESCRIPTION: SEQ ID NO: 1:
; US-09-383-630-1

```

```

alignment_scores:
  Quality: 301.50      Length: 791
  Ratio: 0.908         Gaps: 37
  Percent Similarity: 41.972      Percent Identity: 21.745

alignment_block:
US-10-044-807-2 x US-09-383-630-1 ..

Align seg 1/1 to: US-09-383-630-1 from: 1 to: 5993

1108 IlePheArgSerHisLeuGluHisGlnAspThrLeuLeuLysProSerG1 1124
      :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
2243 GTGTTGGGGGCTGGCGGGGACACCCCAACCCCGCGGGCTGCTGC 2292

1124 uArgArgThrSerProValThrLeuSerProHisLysHisValSerGlyP 1141
      :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
2293 GCGCGGGACACCCAGTTCAGTCACTGCTGTCGACGCGGGGGTGGCGGAG 2342

1141 heSerSerSerLeuArgThrSerSerThrGlyAspAlaGlyGly... 1156
      :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
2343 CAGCTGGCGCGCTGCATCCACTCCGGCGGGGGACTCAGTGGTGGCGG 2392

1157 ..... 1157
      :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
2393 CCGGCCACTGGGACAGAGACACCTGGAAGCGCGCGGACGAGACGGAG 2442

1157 rArg ..... 1160
      :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
2443 CCGCGCTGGTGGTGGCTTACGCGCGGATCCCGCGGCTCGACGA 2492

1161 ..... 1171
      :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
2493 ATTGATATCAAGCTTGCATCCCGCGCGCTGCTGAGGACGCGCGG 2542

1171 eAlaAlaGlnLysSerAlaSerGluValValThrHisLeuGlyGln 1187
      :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
2543 CCGCGCGCCCGCATGGCGCGCTGCTGCGCCCTCGCTCGCGG 2592

1188 ThrValAlaLeuAlaSerGlyThrLeuSerValLeuLeuHisCysGlu... 1203
      :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
2593 GCGGTGGCCATCGTGGCGCGGCTCTCGGAGTCTCTGGGACGAGCA 2642

1203 ..... 1203
      :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
2643 GCGGTGCTGGGCGAGCGGAGAGTCCCGGCGCCAGACCGCGGACG 2692

1204 ..... 1205
      :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
2693 AGGAGCAGTGTGTCCTGGCGAGCGGGAGTCTGTGGAGCTGAGCTGCTCC 2742

```

```

1206 .....GlyHisProArgProThrIleSerTrpAlaArgAsnGlyG1 1219
      :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
2743 CCGCCCGGGGTGGTCCCATGGGCGCCACTGTCTGGGTCAAGGAIGGCAC 2792

1219 uGluValGlnPheSerAspArgIleLeuLeuGlnProAspAspSerLeuG 1236
      :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
2793 AGGCTGTGTCCTCGGAGCGTGTCTGGTGGGCGCC...CAGCGGCTGC 2839

1236 lIleLeuAlaProValGluAlaAspValGlyPheTyrThrCys.AsuAl 1252
      :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
2840 AGTGCTGATGCTCTCCACGAGGACTCCGGGGCTACAGCTGCGGCGAG 2889

1252 aThrAsnAlaLeuGlyTyrAspSerValSerIleAla.....ValThrL 1267
      :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
2890 CGGCTCAGCAGCGCTACTGTGCCTTCACTAGTGTGGGGTGACAGAGCG 2939

1267 euAlaGlyLysProLeuValLysThrSerArgMetThrValIleAsnThr 1283
      :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
2940 TCCATCTCGGGAGATGACGAAGACGGGAG...ACGAGCTGAGGACA 2986

1284 GluLysProAlaValThrValAspIleGlySerThrIleLysThr..... 1298
      :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
2987 CAGGT.....GTGACACAGGGCGCCCTTACTGACACAGCGCC 3023

1299 .....ValGlnGlyValAsn.....ValT 1305
      :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
3024 CGAGCGGATGGACAAGAGCTGCTGGCGTGCAGCGCGCCCAACACCGTCC 3073

1305 hrIleAsnCysGlnValAlaGlyValProGluAlaGluValThrTrpPhe 1321
      :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
3074 GCTTCGCTGCGCCCGCGCTGCAACCCACTCCCTCCATCTCTCTGGCTG 3123

1322 ArgAsnLysSerLysLeuGlySerProHisHisLeu..... 1333
      :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
3124 AAGACGCGCAGGAGTTCCGCGCGAGCACCGCATTTGGAGGCATCAAGCT 3173

1334 .....HisGluGlySerLeuLeuLeuThrAsnValSerSerSerAspG 1348
      :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
3174 GGGCATCAGCAGTGGAGCTGTGATGAAAGCGTGTGCTCGGCTCGGACC 3223

1348 lGlnGlyTyrSerCysArgAlaAlaAsnLeuHisGlyGluLeuThrGlu 1364
      :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
3224 GGGCAACTACACCTCGCTGCGGAGACAAAGTTTGGCAGCATCCGGCAG 3273

1365 SerThrGlnLeuLeuLeuLeuAsp...ProProGlnValProThrGlnLe 1380
      :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
3274 ACGTACAGCTGAGCTGCTGGAGCGCTCCCGGCAC..... 3309

1380 uGluAspIleArgAlaLeuLeuAlaAlaThrGlyProAsnLeuProSerV 1397
      :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
3310 .....CGGCCCCATCTCGAGGCG.....GGGTGCGCGGCCA 3340

1397 allLeuThrSerProLeuGlyThrGlnLeuValLeuAspProGlyAsnSer 1413
      :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
3341 ACCAGACGCGGTGCTGGCAGCAGCTGGAGTTCAC..... 3378

1414 AlaLeuLeuGlyCysProIleLysGlyHisProValProAsnIleThrTr 1430
      :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
3379 .....TGCAAGGTGTACAGTGACGACGACAGCCCGCCACATCCAGTG 3416

1430 pPheHis.....GlyGlyGlnP 1436
      :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
3417 GCTCAAGCAGCTGGAGGTGAACGCGCAGAGGTGGCGCGCGGACGACAC 3466

1436 rIleValThrAlaThrGlyLeuThrHisHisIleLeuAlaAlaGly... 1451
      :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
3467 CCTACGTTACCGTG.....CTCAAGACGCGCGCGCT 3498

1452 .....GlnIleLeuGlnValAlaAsnLeuSerG1 1461
      :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
3499 AACACCACCGACAAAGAGCTAGAGTTCTCTCTTGCACAACTCACTT 3548

```

```
1461 yGlySerGlnGlyGluPheSerLeuAlaGlnAsnGluAlaGlyValL 1478
    : : : ||| ||| : : : : : ||| ||| : : : : : |||
3549 TGAGGACCGCGGGAGTACACTGCTGGCGGCAATCTATTGGGTTT 3598
1478 euMetGlnLysAlaSerLeuValIle.....GlnAsp.TyrTrpTr 1491
    : : : : : ||| ||| : : : : : ||| ||| : : : : : |||
3599 CTCATCACTCTGCTGCTGGTGGTGGTGGTGGTGGTGGTGGTGGT 3648
1491 pSerValAspArgLeuAlaThrCys..... 1499
    | : : : ||| ||| : : : : : ||| ||| : : : : : |||
3649 GAGCTGACGAGCGGGCAGTGTGTATGACGAAATCTCAGCTACAGGGT 3698
1500 ..SerAlaSerCysGlyAsnArgGlyValGlnGlnProArgLeuArgCys 1515
    : : : : : ||| ||| : : : : : ||| ||| : : : : : |||
3699 GGGCTTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 3739
1516 LeuLeuAsnSerThrGluValAsnProAlaHisCysAlaGlyLysValAr 1532
    : : : : : ||| ||| : : : : : ||| ||| : : : : : |||
3740 .....TCTGCCGCTGCGCAGCCCCCAAG 3765
1532 qProAlaValGlnProIleAlaCysAsnArgArgAspCysProSerArgT 1549
    | ||| ||| ||| : : : : : ||| ||| : : : : : |||
3766 AAGGCTGGCTGCCGCCCGCGTGCACAGA..... 3796
1549 rpMetValThrSerTrpSerAlaCysThrArgSerCysGlyGlyVal 1565
    ||| : : : : : ||| ||| : : : : : ||| ||| : : : : : |||
3797 .....TCTCCGCTTCCGCG..... 3811
1566 GlnThrArgArgValThrCysGlnLysLeuLysAlaSerGlyIleSerTh 1582
3811 ..... 3811
1582 rProValSerAsnAspMetCysThrGlnValAlaLysArgPro..... 1596
    ||| : : : : : ||| ||| : : : : : ||| ||| : : : : : |||
3812 .....TCAAGCGACAGTGTCTCCCTGGAGTCCAAACGCTCCATGAGTC 3854
1597 .....ValAspThrGlnAlaCysAsnGlnGlnLeuCysVal 1608
    : : : : : ||| ||| : : : : : ||| ||| : : : : : |||
3855 CAACACACCACTGTCGCGATCCAGCGATGCTCCTCAGGGAGGGCCCCA 3904
1609 GluTrpAlaPheSerSerTrpGlyGlnCys.....AsnGlyPr 1621
    : : : : : ||| ||| : : : : : ||| ||| : : : : : |||
3905 CGTGTGCCAATGCTCCGAGCTCGAGCTGCTGCCGCCGCCCAATGGGAG 3954
1621 oCysIleGlyPro.....H 1626
    ||| : : : : : ||| ||| : : : : : ||| ||| : : : : : |||
3955 CTGCTCGGGCGCGCTGACCTGGGCAAGCCCTTGGGAGGGCTGCTT 4004
1626 isLeuAlaValGlnHisArgGlnValPheCysGlnThrArgAspGly... 1641
    : : : : : ||| ||| : : : : : ||| ||| : : : : : |||
4005 CGGCCAGGTGGTCATGGCGGAGGCCATCGCATTTGACAGGACCGGGCGG 4054
1642 .....IleThrLeuPro...Se 1646
    : : : : : ||| ||| : : : : : ||| ||| : : : : : |||
4055 CCAAGCCTGTACCGCTAGCGGTGAAGATGCTGAAGACCATGCCACTGAC 4104
1646 rGluGlnCys.....SerA 1651
    : : : : : ||| ||| : : : : : ||| ||| : : : : : |||
4105 AAGGACCTGTCGAGCTGCTGAGATGGAGATGATGAAGATGATCGG 4154
1651 laLeuProArgProValSerThrGlnAsnCysTrpSerGluAla..... 1665
    : : : : : ||| ||| : : : : : ||| ||| : : : : : |||
4155 GAAACACAAAACATCATCAACC.....TGCTGGCGCTGCACGACG 4198
1666 .....CysSerValHisTrpArgValSerLeuTrpThrLeuCysTh 1679
    ||| : : : : : ||| ||| : : : : : ||| ||| : : : : : |||
4199 GCGGCCCTGTACCTGCTGTGGTGGATAGCGGCCCAAGGGTAACCTGCGG 4248
1679 rAlaThrCysGlyAsnTyrGlyPheGlnSerArgArgValGluCysValH 1696
    : : : : : ||| ||| : : : : : ||| ||| : : : : : |||
4249 GAGTTTCTGCGGGCGGGCGGCCCGCGCTGGACTACTCTCTCGACAC 4298
1696 isAla...ArgThrAsnLysAlaValProGluHisLeuCysSerTrpGly 1711
```

```
111 ||| ||| : : : : : ||| ||| : : : : : |||
4299 CTGAAGCCGCCCGGAGGAGCAGCTCACCTCAAGG.....ACCTGGTGT 4342

1712 ProArgProAlaAsnTrp 1717
111 ||| ||| : : : : : ||| ||| : : : : : |||
4343 CCTGTGCTACCAAGGTGG 4360

seq_name: /cgn2_6/ptodata/2/ina/6B_COMB.seq:US-09-383-630-2

seq_documentation_block:
  Sequence 2, Application US/09383630A
  Patent No. 6265632
  GENERAL INFORMATION:
  APPLICANT: Avner Yavon et al.
  TITLE OF INVENTION: ANIMAL MODEL FOR FIBROBLAST GROWTH
  FACTOR RECEPTOR ASSOCIATED
  CHONDRODYSPLASIA
  NUMBER OF SEQUENCES: 18
  CORRESPONDENCE ADDRESS:
  ADDRESSEE: Mark M. Friedman c/o Anthony Castorina
  STREET: 2001 Jefferson Davis Highway, Suite 207
  CITY: Arlington
  STATE: Virginia
  COUNTRY: United States of America
  ZIP: 22202
  COMPUTER READABLE FORM:
  MEDIUM TYPE: 1.44 megabyte, 3.5" microdisk
  COMPUTER: Twinhead* Slimnote-890TX
  OPERATING SYSTEM: MS DOS version 6.2,
  Windows version 3.11
  SOFTWARE: Word for Windows version 2.0 converted
  to an ASCII file
  CURRENT APPLICATION DATA:
  APPLICATION NUMBER: US/09/383,630A
  FILING DATE: 26-Aug-1999
  CLASSIFICATION: <Unknown>
  PRIOR APPLICATION DATA:
  APPLICATION NUMBER: <Unknown>
  FILING DATE: <Unknown>
  ATTORNEY/AGENT INFORMATION:
  NAME: Friedman, Mark M.
  REGISTRATION NUMBER: 33,883
  REFERENCE/DOCKET NUMBER: 1402/2
  TELECOMMUNICATION INFORMATION:
  TELEPHONE: 972-3-5625553
  TELEFAX: 972-3-5625554
  TELEX: <Unknown>
  INFORMATION FOR SEQ ID NO: 2:
  SEQUENCE CHARACTERISTICS:
  LENGTH: 5993
  TYPE: nucleic acid
  STRANDEDNESS: double
  TOPOLOGY: linear
  SEQUENCE DESCRIPTION: SEQ ID NO: 2:
  US-09-383-630-2

alignment_scores:
  Quality: 301.50 Length: 791
  Ratio: 0.908 Gaps: 37
  Percent Similarity: 41.972 Percent Identity: 21.745

alignment_block:
  US-10-044-807-2 x US-09-383-630-2 ..
  Align seg 1/1 to: US-09-383-630-2 from: 1 to: 5993

1108 IlePheArgSerHisLeuGluHisGlnAspThrLeuLysProSerG 1124
: : : : : ||| ||| : : : : : ||| ||| : : : : : |||
2243 GTGTTGGGGCTGGCGGAGCACCCCCAACCCCGCGCGGCTGCTGC 2292
1124 uArgArgThrSerProValThrLeuSerProHisLysHisValSerGly 1141
||||| : : : : : ||| ||| : : : : : |||
```

[illegible][illegible]

```

1609 GluTrpAlaPheSerSerTrpGlyGlnCys.....AsnGlyPr 1621
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
3905 CCCTGCCAATGTCCTCCGAGCTCGAGCTGCTGCGACGCCAATGGGAG 3954
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
1621 cCysIleGlyPro.....H 1626
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
3955 CTGTCTCGGGCCCGGTGACCTGCGCAAGCCCTTGGGAGGGTGCCT 4004
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
1626 IsLeuAlaValGlnHisArgGlnValPheCysGlnThrArgAspGly... 1641
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
4005 CGGCCAGGTGGTCTGCGGGAGGCATCGCATTCACAAGGACCGGGCG 4054
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
1642 ..... 1646
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
4055 CCAAGCCTGTACCTAGCCGTGAAGATGCTGAAAGACGATGCCACTGAC 4104
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
1646 rGluGlnCys.....SerA 1651
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
4105 AAGGACCTGTGCGGACCTGCTGCTGAGATGGAGATGATGAAGATCGG 4154
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
1651 laLeuProArgProValSerThrGlnAsnCysTrpSerGluAla..... 1665
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
4155 GAACACAAAACATCATCAACC.....TGCTGGCGCCTGCAGCAGG 4198
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
1666 .....CysSerValHisTrpArgValSerLeuTrpThrLeuCysTh 1679
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
4199 GCGGGCCCTGTACGTGCTGTGAGTAGCGCGGCAAGGGTAACCTGCGG 4248
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
1679 rAlaThrCysGlyAsnTyrGlyPheGlnSerArgValGluCysValH 1696
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
4249 GAGTTTCTGGGGCGCGCGCGCGCGCGCGCTGGACTACTCTTCGACAC 4298
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
1696 isAla....ArgThrAsnLysAlaValProGluHisLeuCysSerTrpGly 1711
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
4299 CTGCAAGCCCGCGGAGGAGGAGCTCACTTCAGG.....ACCTGGTGT 4342
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
1712 ProArgProAlaAsnTrp 1717
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
4343 CCTGTGCTACCAAGGTGG 4360

```

seq_name: /cgn2_6/ptodata/2/ina/5B_COMB.seq:US-08-427-497E-3

```

seq_documentation_block:
; Sequence 3, Application US/08427497E
; Patent No. 5969124
; GENERAL INFORMATION:
; APPLICANT: Lemmon, Vance
; TITLE OF INVENTION: A Method for Characterizing the
; TITLE OF INVENTION: Nucleotide Sequence of L1CAM and
; Patent No. 5969124
; TITLE OF INVENTION: the Nucleotide Sequence
; TITLE OF INVENTION: Characterized Thereby
; NUMBER OF SEQUENCES: 44
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fay, Sharpe, Beall, Fagan,
; ADDRESSEE: Minnich & McKee
; STREET: 1100 Superior Avenue
; STREET: Suite 700
; CITY: Cleveland
; STATE: Ohio
; COUNTRY: U.S.A.
; ZIP: 44114-2518
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 720 Kb
; MEDIUM TYPE: storable
; COMPUTER: Compaq Prolinea 5100e
; OPERATING SYSTEM: DOS 5.0
; SOFTWARE: ASCII
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08427,497E
; FILING DATE: April 24, 1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:

```

```

; APPLICATION NUMBER: 07/904,991
; FILING DATE: June 26, 1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Minnich, Richard J.
; REGISTRATION NUMBER: 24,175
; REFERENCE/DOCKET NUMBER: CWR 2 149-3-1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (216) 861-5582
; TELEFAX: (216) 241-1666
; TELEX: (216) 980162
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3189
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: nucleic acids
; HYPOTHETICAL: irrelevant
; ANTI-SENSE: no
; ORIGINAL SOURCE:
; ORGANISM: homo sapiens
; INDIVIDUAL ISOLATE: 17-18 week fetus
; IMMEDIATE SOURCE:
; LIBRARY: Stratagene cDNA Library 936206
; CLONE: 3.1
; PUBLICATION INFORMATION:
; AUTHORS: Blavin, Mary Louise
; AUTHORS: Lemmon, Vance
; TITLE: Molecular structure and functional testing of
; TITLE: human L1CAM: an interspecies comparison.
; JOURNAL: GENOMICS
; VOLUME: 11
; ISSUE:
; PAGES: 416-423
; DATE: 1991
; RELEVANT RESIDUES IN SEQ ID NO: 548 to 3736
; US-08-427-497E-3

```

alignment_scores:

```

Quality: 287.00 Length: 522
Ratio: 1.206 Gaps: 21
Percent Similarity: 45.594 Percent Identity: 23.755

```

alignment_block:

```

US-10-044-807-2 x US-08-427-497E-3 ..
Align seg 1/1 to: US-08-427-497E-3 from: 1 to: 3189

1135 HisLysHisValSerGlyPheSerSer.....Se 1144
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
81 CAGCCCACTTCCGAGCACCAGGACCATTCATTCAGAGGAACCCATTGA 130

1144 rLeuArgThrSerSerThrGlyAspAlaGlyGlySerArgArgProH 1161
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
131 CCTCCGGGTCAAGGCCACC.....AACAGCATGATTC 162

1161 isArgLysProThrIleLeuArgLysIleSerAlaGlnGlnLeuSer 1177
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
163 ACAGGAAGCGCGCCTGCTCTCCCAACCAAC.....TCC 197

1178 AlaSerGluValValThrHisLeuGlyGlnThrValAlaLeuAlaSerG 1194
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
198 AGCAGCCACTGCTGGCCTTGCAGGGGAG..... 227

1194 yThrLeuSerValLeuHisCysGluAlaIleGlyHisProArgProT 1211
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
228 .....CCATTGGTCTCGAGTGCATCCCGAGGCGCTTCCACGCCCA 270

1211 hrIleSerTrpAlaArgAsnGlyGluGluValGlnPheSerAspArgIle 1227
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
271 CCATCAATGGTGGCGCCCGCAGTGGCCCATGCCA...GCTGACCGTGT 317

```

wed Jul 24 11:35:44 2002

```
1228 LeuLeuGln...ProAspAspSerLeuGlnIleLeuAlaProValGluAl 1243
||| : : : : : ||| : : : : :
318 ACCTACCAACCAACCAAGACCTGTCAGCTGCTGAAGTGGCGAGGA 367
||| : : : : : ||| : : : : :
1243 aAspValGlyPheTyrThrCysAsnAlaThrAsnAlaLeuGlyTyrAspS 1260
||| : : : : : ||| : : : : :
368 GGATGATGCGAGTACCGCTGCTGCGCGAGAACTCACTGGCGAGTGCC 417
||| : : : : : ||| : : : : :
1260 erValSerIleAlaValThrLeuAlaGlyLysProLeuValLysThrSer 1276
||| : : : : : ||| : : : : :
418 GGCATGCGTACTATGACCGTGGAGCGCGCCGACTGG..... 458
1277 ArgMetThrValIleAsnThrGluLysProAlaValThrValAspIleG 1293
||| : : : : : ||| : : : : :
459 .....CTGCACAAGCC..... 470
1293 ySerThrIleLysThrValGlnValAsnValThrIleAsnCysGlnV 1310
||| : : : : : ||| : : : : :
471 .CAGAGCCATCTATATGGCCAGGAGAGACTGCCCGCTGGAGTGCACAG 519
||| : : : : : ||| : : : : :
1310 aAlaGlyValProGluAlaGluValThrTrpPheArgAsnLysSerLys 1326
||| : : : : : ||| : : : : :
520 TCCAGGGCAGGCGCCCAACCAAGAGTCCCTGGAGATCAAC..... 560
1327 LeuGlySerPro.....HisHisLeuHi 1334
||| : : : : : ||| : : : : :
561 ...GGGATCCCTGTGGAGGAGTGGCCAAAGACCAGAAGTACCGGATCA 607
||| : : : : : ||| : : : : :
1334 sGluGlySerLeuLeuLeuThrAsnValSerSerAspGlnLysLeu 1351
||| : : : : : ||| : : : : :
608 GCGTGGCGCTGATCTGACACAGTGCAGCCAGTGCACAAATGTGTGA 657
||| : : : : : ||| : : : : :
1351 yrSerCysArgAlaAlaAsnLeuHisGlyGluLeuThrGluSerThrGln 1367
||| : : : : : ||| : : : : :
658 CCCAATGTGAGGCGCCCAACCGCCGCGCTCTTCTGCGCCCAATGCGTAC 707
||| : : : : : ||| : : : : :
1368 LeuLeuLeuLeuAspProGlnValProThrGlnLeuGluAspIleAr 1384
||| : : : : : ||| : : : : :
708 ATCTACGTGTGC.....CAGCTGCCAGCCAAAG..... 734
1384 gAlaLeuLeuAlaAlaThrGlyProAsnLeuProSerValLeuThrSerP 1401
||| : : : : : ||| : : : : :
735 .....ATCCTGACTGCGG 747
1401 roLeuGlyThrGlnLeuValLeuAspProGlyAsnSerAlaLeuLeuGly 1417
||| : : : : : ||| : : : : :
748 ACAATCAGACCTATAGCTGTCCAG...GGCAGCAGTGCCTACCTCTTG 794
||| : : : : : ||| : : : : :
1418 CysProIleLysGlyHisProValProAsnIleThrTrpPhe...HisG 1433
||| : : : : : ||| : : : : :
795 TCCAGGCGCTTCGGAGCGCCTGTGCCAGTGTTCAGTGGCTGGAGGAGGA 844
||| : : : : : ||| : : : : :
1433 yGlyGlnProIleValThrAlaThrGlyLeuThrHisHisIleLeuAla 1450
||| : : : : : ||| : : : : :
845 TGGGACAACAGTGTCTCAGGACGACGCTTCTCCCTAT.....G 885
||| : : : : : ||| : : : : :
1450 laGlyGlnIleLeuGlnValAlaAsnLeuSerGlySerGlnGlyGlu 1466
||| : : : : : ||| : : : : :
886 CCAATGGGACCTGGGCATTCAGACCTCCAGCCCAATCACACCGGACCC 935
||| : : : : : ||| : : : : :
1467 PheSerCysLeuAlaGlnAsnGluAlaGlyValLeuMetGlnLysAlase 1483
||| : : : : : ||| : : : : :
936 TACTTCTGCTGCTGCCAATGACCAACAACAAATGTACCATCATGCGCTAA 985
||| : : : : : ||| : : : : :
1483 rLeuValIleGlnAspTyrTrpSerValAspArgLeuAlaThrCysS 1500
||| : : : : : ||| : : : : :
986 CTGAGGTTAAAGAT..... 1001
1500 erAlaSerCysGlyAsnArgGlyValGlnGlnProArgLeuArgCysLeu 1516
||| : : : : : ||| : : : : :
1002 .....GCAACTCAGATCACTCAGGGCGCCCGCAGCAATCGAG 1040
||| : : : : : ||| : : : : :
1517 LeuAsnSerThrGluValAsnProAlaHisCysAlaGlyLysValArgPr 1533
```

seq_name: /cn2_5/ptodata/2/ina/5A_COMB.seq.US-07-862-021B-11

```
seq_documentation_block:
; Sequence 11, Application US/07862021B
; Patent No. 5279966
; GENERAL INFORMATION:
; APPLICANT: Jessell, Thomas M
; APPLICANT: Klar, Avihu
; TITLE OF INVENTION: CLONING, EXPRESSION AND USES OF A
; TITLE OF INVENTION: NOVEL SECRETED PROTEIN, F-SPONDIN
; NUMBER OF SEQUENCES: 20
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Cooper & Dunham
; STREET: 30 Rockefeller Plaza
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10112
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/862,021B
; FILING DATE: 19920405
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: White, John P
; REGISTRATION NUMBER: 28,678
; REFERENCE/DOCKET NUMBER: 40028
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 977-9550
; TELEFAX: (212) 664-0525
; TELEX: 422523 COOP UI
; INFORMATION FOR SEQ ID NO: 11:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3226 base pairs
; TYPE: NUCLEIC ACID
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
```



```

;
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 136..2543
US-07-862-021B-11

```

```

alignment_scores:
  Quality: 277.50      Length: 628
  Ratio: 1.110        Gaps: 39
  Percent Similarity: 39.809  Percent Identity: 21.975

alignment_block:
US-10-044-807-2 x US-07-862-021B-11
Align seg 1/1 to: US-07-862-021B-11 from: 1 to: 3226

278 LysIleArgAsnSerGlySerAlaAspSerThrValGlnPheIlePheTy 294
|||||
1243 AAGATTAGACCCTTACAGCTTAGATCACCT...CAGAGTCCATTTTA 1289
|||||
294 rGlnProIleIleHisArgTrrArgGluThrAspPheProCysSerA 311
|:|:|
1290 TGATCCAGAA..... 1299

311 laThrCysGlyGlyGlyTyrGlnLeuThrSerAlaGluCysTyrAspLeu 327
|||||
1300 .....GGAGATCTATCAAGCTTGAGCC..... 1323

328 ArgSerAsnArgValValAlaAsp.....GlnTyrCy 338
|||||
1324 .....AGATCGTCTTGAAGAATTGCACCAAGGGGACCACTG 1364

338 sHisTyrTrpProGluAsnIle.....LysProL 348
|:|:|
1365 CAATTCTGCTGATACATAGATGATGATTGTGGCAGACCTAGCACCG 1414

348 ysProLysLeuGlnGluCysAsnLeuAspProCysProAlaSerAspGly 364
|:|:|
1415 AAGAAAGAGAGAGATGATACCTCGAGACCTGCATATATCAAC... 1461

365 TyrLysGlnIleMetProTyrAspLeuTyrHisProLeuProArgTrrpGl 381
|:|:|
1462 .....TGG... 1464

381 uAlaThrProTrrpThrAlaCysSerSer...CysGlyGlyGlyIleg 397
|:|:|
1465 .....TCCCTCGTGCAGCTGCTCTCTACCTGTGAGAAGGCAAGA 1510

397 lnSerArgAlaValSerCysValGluGluAspIleGlnGlyHisValThr 413
|:|:|
1511 GGATGAGG..... 1518

414 SerValGluGluTrpLysCysMetTyrThrProLysMetProIleAlaGl 430
|:|:|
1519 .....CAGAGATGCTTTAAAGCTCAGCTGAGCTTCAGTGT 1553

430 nProCysAsnIlePheAspCysProLysTrrpLeuAlaGlnGluTrpSerp 447
|:|:|
1554 GCCC.....TGTCCTGAT.....ACCCAAGATTTCAGC 1582

447 roCysThrValThrCysGlyGlnGlyLeuArgTyrArgValValLeuCys 463
|:|:|
1583 CARGC.....ATGGGTCCAGGC.....TGC 1602

464 IleAspHisArgGlyMetHisThrGlyGlyCysSerProLysThrLysPr 480
|:|:|
1603 AGTGATGAAGATGGT..... 1617

480 oHisIleLysGluGluCysIleValProThrProCysTyrLysProLysG 497
|:|:|
1618 .....TCAACTTGCAATGATGCT..... 1635

497 luLysLeuProValGluAlaLysLeuProTrrpPheLysGlnAlaGlnGlu 513
|:|:|

```

```

1636 .....GACTGGATT..... 1644
514 LeuGluGluGlyAlaAlaValSerGluGluProSerPheIleProGluAl 530
|:|:|
1645 .....AC 1646
530 aTrpSerAlaCysThrValThrCysGlyValGlyThrGlnValArgIlev 547
|:|:|
1647 ATGGTCCCTGTAGTGTTCCTGTGGAATGGGAACGCGATCTAGA... 1692
547 alArgCysGlnValLeuLeuSerPheSerGlnSerValAlaLaAspLeuPro 563
|:|:|
1693 .....GAGAGATATGTAAAGCAATTCCTCC 1716
564 IleAsp.....GluCysGluGlyProLysProAlaSerGlnArgAlaCy 578
|:|:|
1717 GAAGATGGCTCTATGTGCAAGTGCCTACTGAAAGAACTGAGAAATGTAT 1766
578 sTyrAlaGlyProCysSerGlyGlu.....IleProGluPheAsnP 592
|:|:|
1767 TGTAAATGAGGATGCTCCCTAGCAGCTGCCTTGTACCGAATGGGA. 1815
592 roAspGluThrAspGlyLeuPheGlyGlyLeuGlnAspPheAspGluLeu 608
|:|:|
1815 ..... 1815
609 TyrAspTrpGluTyrGluGlyPheThrLysCysSerGluSerCysGlyG 625
|:|:|
1816 ...GAGTGGGAT.....GAATGCAGTGTAGCTGTGGCAC 1847
625 yGlyValGlnGluAlaValValSerCysLeuAsnLysGlnThrArgGluP 642
|:|:|
1848 AGGAATGAAAGGCGA.....CACAGATGATCAAGATGACT.....C 1885
642 roAlaGluGluAsnLeuCys.....ValThrSerArgArgProProGln 656
|:|:|
1886 CTGCTGATGATCTATGTGCAAGGCGAAGACTACAGAGCGACAGAAATGC 1935
657 LeuLeuLysSerCysAsnLeuAspProCys...ProAlaArgTrrpGluI 672
|:|:|
1936 ATGATGCCCCGAATGCCACTACTATTCCTTCCTTCATCCCATGG... 1980
672 eGlyLysTrpSerProCysSerLeuThrCysGlyValGlyLeuGlnThrA 689
|:|:|
1981 .TCGAAATGGAGCGACTGCAGCGTGCATGTGGGAAGGGAATGCGAACCC 2029
689 rgAspValPheCysSerHisLeuLeuSerArgGluMetAsnGlnThrVal 705
|:|:|
2030 GG.....CAAAGGATGCTGAATTCGCACCT 2055
706 IleLeuAlaAspGluLeuCysArgGlnProLysProSerThrValGlnAl 722
|:|:|
2056 GAGCTTGGAGAC.....TGCAATGAG...GAAGTGGAGCAAGCAGAGAA 2096
722 aCysAsnArgPheAsnCysPro.....ProIleTrrpTyrp 734
|:|:|
2097 ATGATGCTACCTGAATGCCCATGCTGAGCTAACGGAGTGG... 2142
734 roAlaGlnTrpGlnProCysSerArgThrCysGlyGlyValGlnLys 750
|:|:|
2143 .TCCCACTGTGCGAGTGCATATACCTCTCTGTGGGAAGGCGCACATGATC 2190
751 ArgGluValLeuCysLysGlnArgMetAlaAspGlySerPheLeuGluLe 767
|:|:|
2191 AGAACAGAAGATGATCAAAATAGAACCACAGTCTTGGAGGA...ACAGCATG 2237
767 uProGluThrPhe.....Cys..... 772
|:|:|
2238 CCCAGAAACTGTCCAGCTACTAAATGTCAGTAAGGAATGCTGAGAG 2287
773 .....SerAlaSerLysProAla 778
|:|:|

```



```
547 aArgCysGlnValLeuLeuSerPheSerGlnSerValAlaAspLeuPro 563
1693 .....GAGAGATATGTAAGCAATTCCTCC 1716
564 IleAsp.....GluCysGluGlyProLysProAlaSerGlnArgAlaCys 578
1717 GAAGATGGCTCTATGTGCAAGAGCTCTATCTGAAGAACTGAGAAATGTAT 1766
578 sTyrAlaGlyProCysSerGlyGlu.....IleProGluPheAsn 592
1767 TGTAAATGAGGAATGCTCCCTAGCAGCTGCTGTGTACCGAATGGGA. 1815
592 roAspGluThrAspGlyLeuPheGlyGlyLeuGlnAspPheAspGluLeu 608
1815 ..... 1815
609 TyrAspTrpGluTyrGluGlyPheThrLysCysSerGluSerCysGlyGly 625
1816 ...GAGTGGGAT.....GAATGCACTGCTAGCTGGGCAC 1847
625 yGlyValGlnGluAlaValSerCysLeuAsnLysGlnThrArgGluP 642
1848 AGGAATGAAAGGGGA.....CACAGAATGATCAAGATGACT.....C 1885
642 roAlaGluGluAsnLeuCys.....ValThrSerArgArgProGln 656
1886 CTGCTGATGGATCTATGTCAGGCGAGAACTACAGAGCGAGAGAAATGC 1935
657 LeuLeuLysSerCysAsnLeuAspProCys...ProAlaArgTrpGlu 672
1936 ATGATGCCGAATGCCATCTATCTCCCTGCCTCTCTCCCATGG..... 1980
672 eGlyLysTrpSerProCysSerLeuThrCysGlyValGlyLeuGlnThr 689
1981 .TCGAATGGAGCGACTGCACGTGACATGTGGGAAGGAATGCAAGCC 2029
689 rgAspValPheCysSerHisLeuLeuSerArgGluMetAsnGluThrVal 705
2030 GG.....CAAGGATGCTGAAATCTGCAGCT 2055
706 IleLeuAlaAspGluLeuCysArgGlnProLysProSerThrValGlnAl 722
2056 GAGCTTGGAGAC.....TGCAATGAG...GAAGTGGAGCAGCAGAGAA 2096
722 aCysAsnArgPheAsnCysPro.....ProAlaTrpTyrP 734
2097 ATGATGCTACCTGAATGCCCATTTGACTGTGAGCTAACGGAGTGG.... 2142
734 roAlaGlnTrpGlnProCysSerArgThrCysGlyGlyValGlnLys 750
2143 ..TCCCAAGTGGTCCGAGTGCATACCTCTCTGTGGGAAGGGCCACATGATC 2190
751 ArgGluValLeuCysLysGlnArgMetAlaAspGlySerPheLeuGluLe 767
2191 AGAACAAGATGATCAAAATAGAACACACAGTTTGGAGGA...ACAGCATG 2237
767 uProGluThrPhe.....Cys..... 772
2238 CCCAGAAACGTCCCAACGTACTAAATGTCGAGTAAGGAATGCTGAGAG 2287
773 .....SerAlaSerLysProAla 778
2288 GCCCAGGTATGGAAGAGCGGTGGAAGAGCGCCGGGAGAAAGAAAGA 2337
779 CysGlnGlnAlaCysLysLysAspAsp.....CysPr 789
2338 AGTGAACAAGCAAAAAAATATTGATATGATGCAATATCCAGTTTGTAG 2387
789 o....SerGluTrpLeuLeuSerAspTrpThrGluCysSerThrSerCysG 805
2388 GCTGAAACCATG.....ACTGCTGGACAGAAATGTTCTACACTCTGTC 2431
```

```
805 lyGluGlyThrGlnThrArgSerAlaIleCysArgLysMetLeuLysThr 821
2432 GAGGTGAATTCAGGACCGCTACATGATGTTAAAGAAAGAGTCCAAAAGC 2481
822 GlyLeuSerThrValValAsnSerThrLeuCysProProLeuProPheSe 838
2482 .....ACTCAGTTTACTAGTGCAAAGAC.....AA 2507
838 rSerSerIleArgProCysMetLeuAlaThrCys 849
2508 AAAGGAGCTAAGAGCATGTAACGTTTCATCCTTGT 2541
seq_name: /cgn2_6/ptodata/2/ina/PCTUS_COMB.seq:PCT-US93-03164-11
seq_documentation_block:
; Sequence 11, Application PC/TUS9303164
; GENERAL INFORMATION:
; APPLICANT: Jessell, Thomas M
; TITLE OF INVENTION: CLONING, EXPRESSION AND USES OF A
; TITLE OF INVENTION: NOVEL SECRETED PROTEIN, F-SPONDIN
; NUMBER OF SEQUENCES: 20
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Cooper & Dunham
; STREET: 30 Rockefeller Plaza
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10112
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US93/03164
; FILING DATE: 19930402
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: White, John P
; REGISTRATION NUMBER: 28,678
; REFERENCE/DOCKET NUMBER: 40028
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 977-9550
; TELEFAX: (212) 664-0525
; TELEX: 422523 COOP UI
; INFORMATION FOR SEQ ID NO: 11:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3226 base pairs
; TYPE: NUCLEIC ACID
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 136..2543
PCT-US93-03164-11
```

```
alignment_scores:
Quality: 277.50      Length: 628
Ratio: 1.110        Gaps: 39
Percent Similarity: 39.809      Percent Identity: 21.975
```

```
alignment_block:
US-10-044-807-2 x PCT-US93-03164-11 ..
```

```
Align seg 1/1 to: PCT-US93-03164-11 from: 1 to: 3226
```

```
278 LysileArgAsnSerGlySerAlaAspSerThrValGlnPheIlePheTy 294
```

1243 AAGATTAGACACATTACAGCTTAGATCACCT...CAGAGTCCATTTTA 1289
294 rGlnProIleIleHisArgTyrTrpArgGluThrAspPheProCysSerA 311
1290 TGATCCAGAA 1299
311 laThrCysGlyGlyGlyTyrGlnLeuThrSerAlaGluCysTyrAspLeu 327
1300GGAGGATCTATCAAGCTTGATGCC 1323
328 ArgSerAsnArgValValAlaAspGlnTyrCys 338
1324AGAGTCTGCTCTGAAGAAGATTGCACGCAAGGGGAGCAGTG 1364
338 shiSTyTyrProGluAsnIleLysProL 348
1365 CAACCTCGTACCTGATACATAGATGATATTGTGCGACACCTAGCACCAG 1414
348 ySProlYsLeuGlnGlnCysAsnLeuAspProCysProAlaSerAspGly 364
1415 AGAAAAAGAAAGATGATACCCCTGACGACTGCATATATCAAC... 1461
365 TyrLysGlnIleMetProTyrAspLeuTyrHisProLeuProArgTyrPgl 381
1462TGG... 1464
381 uAlaThrProTyrThrAlaCysSerSerSer...CysGlyGlyGlyIleG 397
1465TCCCCCTGGTCAGCTGCAGCTCTCTACCTGTGAGAAGGCAAGA 1510
397 lnSerArgAlaValSerCysValGluGluAspIleGlnGlyHisValThr 413
1511 GGATGAGG 1518
414 SerValGluGluThrLysCysMetTyrThrProLysMetProIleAlaG 430
1519CAGAGATGCTTAAGCTCAGCTGACCTCAGCTGT 1553
430 nProCysAsnIlePheAspCysProLysTrpLeuAlaGlnGluTrpSerP 447
1554 GCCCACCCAAGATTTTCAGC 1582
447 roCysThrValThrCysGlyGlnGlyLeuArgTyrArgValValLeuCys 463
1583 CATGCTGC 1602
464 IleAspHisArgGlyMetHisThrGlyGlyCysSerProLysThrLysPr 480
1603 AGTGATGAAGATGT 1617
480 oHisIleLysGluGluCysIleValProThrProCysTyrLysProLysG 497
1618TCAACTTGCATGATGTCT 1635
497 luLysLeuProValGluAlaLysLeuProTyrPheLysGlnAlaGlnGlu 513
1636GACTGGATT 1644
514 LeuGluGluGlyAlaAlaValSerGluGluProSerPheIleProGluAl 530
1645AC 1646
530 aTrpSerAlaCysThrValThrCysGlyValGlyThrGlnValArgIleV 547
1647 ATGTCCTCCCTGTAGTCTTCTGTAATGGGAACCGCATCAGA... 1692
547 alArgCysGlnValLeuLeuSerPheSerGlnSerValAlaAspLeuPro 563
1693GAGAGATGTAAAGCAATTCCTCC 1716
564 IleAsp.....GluCysGluGlyProLysProAlaSerGlnArgAlaCy 578
1111111111 111 111 111 1111111111
1717 GAAGATGGCTCTATGTGCAAAAGTGCCTACTGAAGAACTGAGAAATGTAT 1766
578 styrAlaGlyProCysSerGlyGlu.....IleProGluPheAsnPr 592
1767 TGTAAATGAGGAATGCTCCCTAGCAGCTGCTTCTCACCAGATGGGA. 1815
592 roAspGluThrAspGlyLeuPheGlyLeuGlnAspPheAspGluLeu 608
1815 1815
609 TyrAspTrpGluTyrGluGlyPheThrLysCysSerGluSerCysGlyG 625
1816 ...GAGTGGGATGAATGCATGCTAGCTGTGGCAG 1847
625 yGlyValGlnGluAlaValValSerCysLeuAsnLysGlnThrArgGluP 642
1848 AGCAATGAAAAAGCGGA.....CACAGAATGATCAAGATGACT.....C 1885
642 roAlaGluGluAsnLeuCys.....ValThrSerArgArgProGln 656
1886 CTCTCATGATGATCTATGTCAAAGCGCAGAACTACAGAGCGACAGAAATGC 1935
657 LeuLeuLysSerCysAsnLeuAspProCys...ProAlaArgTyrPglu 672
1936 ATATGCCCGAATGCCATACTATTCTCTGCTTCTATCCCATGG..... 1980
672 eGlyLysTrpSerProCysSerLeuThrCysGlyValGlyLeuGlnThrA 689
1981 .TCTGAATGGAGCGACTGCAGCGTGACATGTGGGAAGGCAATCGCAACC 2029
689 rgAspValPheCysSerHisLeuLeuSerArgGluMetAsnGluThrVal 705
2030 GGCAAAGGATGCTGAAATCTGCAGCT 2055
706 IleLeuAlaAspGluLeuCysArgGlnProLysProSerThrValGlnAl 722
2056 GAGCTTGGAGAC.....TGCAATGAG...GAACCTGGAGCAAGCAGAGAA 2096
722 acYsAsnArgPheAsnCysPro.....ProAlaTrpTyrP 734
2097 ATGCTACTACCTCAATGCCCATCTGACTGTGAGCTAACGGAGTGG... 2142
734 roAlaGlnTrpGlnProCysSerArgThrCysGlyGlyValGlnLys 750
2143 .TCCCCAGTGTGCGAGTGCATACCTCTCTGGGAAGGCCACATCATC 2190
751 ArgGluValLeuCysLysGlnArgMetAlaAspGlySerPheLeuGluLe 767
2191 AGACACAGAAATGCAAAATAGAACCCAGCTTTGGAGGA...ACAGCATG 2237
767 uProGluThrPhe.....Cys..... 772
2238 CCCAGAACTGTCCAACGTACTAAATCTCGAGTAAGGAAATGCTCGAGAG 2287
773SerAlaSerLysProAla 778
2288 GCCCAGGTATGGAAGAGAGCGCTTGGAAAGGCGCCCGGAGAAAAAGAA 2337
779 CysGlnGlnAlaCysLysLysAspAsp.....CysPr 789
2338 ACTGACCAAGCAAAAAAATAATTGATAATGACGCAATATCCAGTTTGTAG 2387
789 o...SerGluTrpLeuLeuSerAspTrpThrGluCysSerThrSerCysG 805
2388 GCTGAAACCATCGACTGCTTGCACAGAAATGTCTACACTCTGTG 2431
805 yGluGlyThrGlnThrArgSerAlaIleCysArgLysMetLeuLysThr 821
2432 GAGGTGGAATTCAGGAGCGCTACATGATGGTAAAGAGAGAGGTCCAAAGC 2481
822 GlyLeuSerThrValValAsnSerThrLeuCysProProLeuProPheSe 838
2482ACTCAGTTTACTAGCTGCAAAAGACAA 2507

```
838 rSerSerIleArgProCysMetLeuAlaThrCys 849
      ::::::::::|||:::
2508 AAAGGAGCTAAGAGCATGTACGTTTCCTTGT 2541
```


OM of: US-10-044-807-2 to: EST:* out_format : pfs

Date: Jul 24, 2002 5:14 AM

About: Results were produced by the GenCore software, version 4.5,
Copyright (c) 1993-2000 CompuGen Ltd.

Command line parameters:

```
-MODEL=frame+p2n.model -DEV=xlh
-Q/cgn2.1/USPTO.spool/US10044807/runat_22072002.153741.13644/app_query.fasta_1.1847
-DB=EST -QWMT=Eastap -SUFFIX=rst -CAPOP=12.000 -GAPEXT=4.000
-MINMATCH=0.100 -LOOPL=0.000 -LOOPEXT=0.500 -FGAPOP=4.500
-QGAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500 -FGAPOP=6.000
-FGAPEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500 -DELOP=6.000
-DELETE=7.000 -START=1 -MATRIX=blosum62 -TRANS=human40.cdi
-LIST=45 -MODALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0
-ALIGN=15 -MODEL=LOCAL -OUTEXT=pfs -NORM=ext -HEAPSIZ=500
-MINLEN=0 -MAXLEN=200000000 -USER=US10044807 @CGNL1.1.4168
-NCPU=6 -ICPU=3 -LONGLOG -DEV_TIMEOUT=120 -WARN_TIMEOUT=30
-NO_MLPHY -WAIT -THREADS=1
```

Search information block:

Query: US-10-044-807-2

Query length: 1762

Database: EST:*

Database sequences: 13736207

Database length: -1841457050

Search time (sec): 2791.330000

score_list:

Sequence	Strd Orig	zScore	EScore	Len	Documentation
gb_est2:BM476141	+ 1614.00	2227.34	8.1e-115	1009	! BM476141 AGNCOURT_6478885 NIH
gb_est2:BE888902	+ 1175.50	1618.86	6.3e-81	850	! BE888902 601513828F1 NIH_MGC_71
gb_est1:BB633755	+ 1102.00	1520.30	2.0e-75	622	! BB633755 BB633755 RIKEN full-le
gb_est1:AI917724	- 1044.00	1440.64	5.4e-71	568	! AI917724 ttlic08.xl NCI_CGAP_G
gb_htc:AK020115	+ 1037.00	1424.31	4.3e-70	979	! AK020115 Mus musculus 12 days e
gb_est1:BB652922	+ 998.00	1374.77	2.5e-67	660	! BB652922 BB652922 RIKEN full-le
gb_est1:BI523618	+ 958.00	1320.62	2.6e-64	581	! BI523618 603051757F1 NIH_MGC_12
gb_est1:AI342006	- 926.00	1319.13	3.1e-64	522	! AI342006 qs98d10.xl NCI_CGAP_G
gb_est1:BB643318	+ 942.00	1296.09	6.0e-63	700	! BB643318 BB643318 RIKEN full-le
gb_est1:BB478282	+ 934.00	1286.44	2.1e-62	619	! BB478282 BB478282 RIKEN full-le
gb_est1:BB620571	+ 932.50	1284.08	2.8e-62	633	! BB620571 BB620571 RIKEN full-le
gb_est1:BB612331	+ 888.00	1222.04	8.0e-59	637	! BB612331 BB612331 RIKEN full-le
gb_est1:AI261611	+ 869.00	1198.30	1.7e-57	509	! AI261611 qz35h01.xl NCI_CGAP_K
gb_est1:BB193444	+ 842.00	1157.88	3.0e-55	643	! BB193444 BB193444 RIKEN full-le
gb_est1:BE628435	+ 839.00	1155.59	4.0e-55	550	! BE628435 uu09f09.Y2 Soares mam
gb_est1:BF058634	- 826.00	1139.94	3.0e-54	449	! BF058634 7k24d09.xl NCI_CGAP_O
gb_est1:BB615232	+ 811.00	1116.50	6.1e-53	555	! BB615232 BB615232 RIKEN full-le
gb_htc:AK020361	+ 775.00	1061.28	7.2e-50	846	! AK020361 Mus musculus adult mal
gb_est2:BF510853	- 765.50	1052.49	2.2e-49	586	! BF510853 UI-H-B14-apm-e-10-0-U
gb_est2:BG750735	- 759.00	1043.03	7.5e-49	606	! BG750735 602706924F1 NIH_MGC_4
gb_est2:N33407	- 750.00	1033.99	2.4e-48	454	! N33407 yv41f11.s1 Soares meland
gb_est2:BF111214	- 713.00	978.25	3.0e-45	644	! BF111214 7n44e07.xl NCI_CGAP_L
gb_est1:AI825646	- 709.00	979.06	2.7e-45	380	! AI825646 wbf5b10.xl NCI_CGAP_G
gb_est1:AA482392	+ 693.50	956.18	5.2e-44	423	! AA482392 zv05h07.Y1 Soares NHHM
gb_est2:BE631939	+ 693.00	951.26	9.7e-44	600	! BE631939 uu09f09.X1 Soares mam
gb_est1:AI750970	+ 681.00	941.67	3.3e-43	419	! AI750970 cn06h02.Y1 Normal Huma
gb_est2:BF45136	- 651.00	932.17	1.9e-40	631	! BF45136 nad21a04.xl NCI_CGAP_L
gb_est1:AA120520	- 607.00	892.17	3.8e-37	539	! AA120520 UI-M-BH1-ana-b-01-0-U
gb_est1:AI459225	- 602.50	824.70	1.1e-36	628	! AI459225 tk11b03.xl NCI_CGAP_L
gb_est1:BB839998	- 593.00	801.79	2.1e-35	442	! BB839998 BB839998 RIKEN full-le
gb_est1:AA120794	- 573.00	787.60	1.3e-34	452	! AA120794 UI-M-BH2.3-any-h-08-0
gb_est2:BF67952	+ 555.00	763.28	2.9e-33	425	! BF67952 CM1-CN0061-201200-673-
gb_est1:AA284278	+ 548.00	757.23	6.2e-33	313	! AA284278 zc39a07.Y1 Soares_sensc
gb_est1:AA284278	+ 525.00	725.64	3.6e-31	302	! AA284278 zc39a07.Y7 Soares_sensc
gb_est1:AA791699	- 461.00	629.64	8.0e-26	534	! AA791699 op95a12.s1 NCI_CGAP_L
gb_est2:W47029	- 457.00	626.49	1.2e-25	437	! W47029 zc39a07.Y1 Soares_sensc
gb_gss:CN503V8J	+ 430.50	579.80	4.8e-23	983	! AT262108 Tetradon nigroviridis
gb_est2:N44869	+ 427.00	592.32	9.6e-24	233	! N44869 yv40f11.xl Soares meland
gb_htc:BC003269	+ 426.50	561.03	5.3e-22	2931	! BC003269 Mus musculus, Similar

gb_est1:AW027573 - 420.50 571.00 1.5e-22 643 ! AW027573 wv74h06.x1 Soares t
gb_est2:BI821835 + 418.50 565.81 2.9e-22 785 ! BI821835 603035722F1 NIH_MGC
gb_est1:AL665815 + 409.50 549.89 2.2e-21 1039 ! AL665815 AL665815 direction
gb_est2:BI733795 + 409.00 553.20 1.4e-21 746 ! BI733795 603351940F1 NIH_MGC
gb_est2:BI852900 + 407.00 548.68 2.6e-21 861 ! BI852900 603379705F1 NIH_GGA

seq_name: gb_est2:BM476141

seq documentation block:

LOCUS BM476141 1009 bp mRNA linear EST 05-FEB-2002
DEFINITION AGNCOURT_6478885 NIH_MGC_88 Homo sapiens cDNA clone IMAGE:5558669
5', mRNA sequence.
ACCESSION BM476141
VERSION BM476141.1 GI:18525183
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 1009)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgabbs@mail.nih.gov
Tissue Procurement: ATCC
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLAM12282 row: a column: 06
High quality sequence stop: 647.

FEATURES
Source

1..1009
Location/Qualifiers
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:5558669"
/clone_lib="NIH_MGC_88"
/tissue_type="duodenal adenocarcinoma, cell line"
/lab_host="DH10B (phage-resistant)"
/notes="Organ: small intestine; Vector: pCMV-SPORT6;
Site.1: NotI; Site.2: SalI; Cloned unidirectionally;
oligo-dt primed. Average insert size 1.767 kb. Library
enriched for full-length clones and constructed by Life
Technologies. Note: this is a NIH_MGC Library."

BASE COUNT 234 a 299 c 268 g 204 t 4 others

alignment_scores:

Quality: 1614.00 Length: 334
Ratio: 5.091 Gaps: 11
Percent Similarity: 94.910 Percent Identity: 92.814

alignment_block:

US-10-044-807-2 x BM476141 ..

Align seg 1/1 to: BM476141 from: 1 to: 1009

270 ProLeuThrAlaAspPheIleValIysIleArgAsnSerGlySerAlaAs 286
|||||

26 CCACCTCACAGCAGATTCATTGTCAAGATTCGTAACCTCGGCTCGCTGA 75
|||||

286 pSerThrValGlnPheIlePheTyrGlnProIleIleHisArgTrpArg 303
|||||

76 CAGTACAGTCCAGTTCATCTCTATCACCCATCATCCGATGAGGAGG 125
|||||

303 LuThrAspPhePheProCysSerAlaThrCysGlyGlyTyrGlnLeu 319
|||||

126 AGACGGATTTCTTCTTCGTCAGCACTGAGGAGGATTATCAGCTG 175
|||||


```

320 ThrSerAlaGluCysTyrAspLeuArgSerAsnArgValValAlaAspG1 336
|||||
176 ACATCGGCTGAGTGCTAGCATCTGAGGAGCAACCGTGTGTTGCTGACCA 225
|||||
336 nTyrCysHisTyrTyrProGluAsnIleLeuProLysProLysLeuGlnG 353
|||||
226 ATACTGTCACTATACCCAGAGACATCAACCCAAACCAAGCTTCAGG 275
|||||
353 luCysAsnLeuAspProCysProAlaSerAspGlyTyrLysGlnIleMet 369
|||||
276 AGTGCAACTTGATCTCTGTGCAGCCAGTCAGGATACAAGCAGATCATG 325
|||||
370 ProTyrAspLeuTyrHisProLeuProArgTyrProGluAlaThrProTyr 386
|||||
326 CCTATGACCTCTACCATCCCTCTCTCGGTGGAGGCCACCCCATGGAC 375
|||||
386 rAlaCysSerSerCysGlyGlyGlyIleGlnSerArgAlaValSerC 403
|||||
376 CGCGTGTCTCTCTCTCGGTGGGGGGGATCCAGAGCCGGCAGTTTCTCT 425
|||||
403 ysValGluGluAspIleGlnGlyHisValThrSerValGluGluThrLys 419
|||||
426 GTGTGGAGGAGACATCCAGGGGATGTCACCTTCAGTGGAAAGTGGAAA 475
|||||
420 CysMetTyrThrProLysMetProIleAlaGlnProCysAsnIlePheAs 436
|||||
476 TGCATGTACACCCCTAGATGCCATCGCAGCGCCCTGCAACACTTTTGA 525
|||||
436 pCysProLysTrpLeuAlaGlnGluTrpSerProCysThrValThrCysG 453
|||||
526 CTGCGCCCTAAATGGTGGACAGAGAGTGTCCTCGGTGCACAGTGACATG 575
|||||
453 lyGlnGlyLeuArgTyrArgValValLeuCysIleAspHisArgGlyMet 469
|||||
576 GCCAGGGCTCAGATACCGTGTGCTCTCTGCATCGACCATCGAGGAATG 625
|||||
470 HisThrGlyGlyCysSerProLysThrLysProHisIleLysGluLucy 486
|||||
626 CACACAGAGGCTGTAGCCCAAAACAAAGCCCAACATAAAAGAGGAATG 675
|||||
486 sIleValProThrProCysTyrLysProLysGluLysLeuProValGluA 503
|||||
676 CATCGTACCCACTCCCTGTCTATAACCAACCAAGAGAACTTCCAGTCGAGG 725
|||||
503 laLysLeuProTrpPheLysGlnAlaGlnGluLeuGluGluGlyAlaAla 519
|||||
726 CCAAGTTGCCATGTTCAAAACAGCTCAAGAGCTAGAAAGAGAGCTGCT 775
|||||
520 ValSerGluGluProSerPheIleProGluAlaTrpSerAlaCysThr.V 536
|||||
776 GTGTACAGAGAGCCCTGTTCTATCCAGAGCCCTGGTGGGCTGCACNAG 825
|||||
536 alThrCysGlyValGlyThrGlnValArg.IleValArgCysGln.ValL 552
|||||
826 TCACCTGTGGTGTGGGACCCAGCTGGGAATAGTCAGTGCACGNGGC 875
|||||
552 euLeuSerPhe.SerGlnSer.ValAlaAspLeu.ProIleAspGluCys 567
|||||
876 TCCGTCTCTCTCTAGNCCCGTGGGTGACCTGGCCCTATTGACCCAGTG 925
|||||
568 ...GluGlyProLys.ProAlaSerGlnArg.AlaCysTyrAla.GlyPr 582
|||||
926 TGGAAAGGGCCCTAGCCAGCATCCAGCTGCGGCTGTGTTATGAGGGGCC 975
|||||
582 oCys...SerGlyGluLeuProGluPheAsnPro 592
|||||
976 ATGCCACGGGGGAAAAATCTCTGAGTCCACCCC 1009
|||||

```

seq_name: gb_est2:BE888902

seq_documentation_block:

LOCUS BE888902 850 bp mRNA linear EST 20-OCT-2000
DEFINITION 601513828F1 NIH_MGC_71 Homo sapiens cDNA clone IMAGE:3915294 5',

ACCESSION VERSION KEYWORDS SOURCE

BE888902
BE888902.1 GI:10345670
EST
human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 850)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: ATCC
CDNA Library Preparation: Life Technologies, Inc.
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LML at:
http://image.llnl.gov
Plate: LLWG9738 row: g column: 07
High quality sequence stop: 742.

FEATURES source

1..850
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:3915294"
/clone_lib="NIH_MGC_71"
/tissue_type="leiomyosarcoma"
/lab_host="DH10B (phage-resistant)"
/note="Organ: uterus; Vector: pCMV-SPORT6; Site_1: NotI;
Site_2: SalI; Cloned unidirectionally. Primer: Oligo dt.
Average insert size 2.1 kb."
BASE COUNT 202 a 219 c 228 g 201 t
ORIGIN

alignment_scores:

Quality: 1175.50 Length: 269
Ratio: 4.798 Gaps: 5
Percent Similarity: 91.078 Percent Identity: 86.617

alignment_block:

US-10-044-807-2 x BE888902 ..

Align seg 1/1 to: BE888902 from: 1 to: 850

1 MetGluCysCysArgArgAlaThrProGlyThrLeuLeuLeuPheLeuAl 17
|||||
44 ATGGAATGCTGCCGTGGGCAACTCCTGCGACACTGCTCTTTCIGGC 93
|||||
17 aPheLeuLeuLeuSerSerArgThrAlaArgSerGluGluAspArgAspG 34
|||||
94 TTTCTGCTCTGAGTTCCAGACCCGACGCTCCGAGGAGGACCGGAGC 143
|||||
34 lyLeuTrpAspAlaTrpGlyProTrpSerGluCysSerArgThrCysGly 50
|||||
144 GCCTATGGATGCTCCGTGGGCCCATGGAGTGAATGCTCAGCACCTGCGGG 193
|||||
51 GlyGlyAlaSerTyrSerLeuArgArgCysLeuSerSerLysSerCysG1 67
|||||
194 GGTGGGCTCTCTACTCTCTGAGGCGCTGCTGAGCAGCAGAGCTGTA 243
|||||
67 uGlyArgAsnIleArgTyrArgThrCysSerAsnValAspCysProProG 84
|||||
244 AGGAAGAAATATCCGATACAGACATGAGTGAATGTGGACTGCCCCACAG 293
|||||
84 luAlaGlyAspPheArgAlaGlnGlnCysSerAlaHisAsnAspValLys 100
|||||
294 AAGCAGGTGATTTCCGAGCTCAGCAATGCTCAGCTCATAATGATGTCAG 343
|||||
101 HisHisGlyGlnPheTyrGluTrpLeuProValSerAsnAspProAspAs 117
|||||


```

29 uGluAspArgAspGlyLeuTrpAspAlaTrpGlyProTrpSerGluCys 46
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
52 GGAAGACCGTCAAGCGCTCTCGGATGCTGGGCCCTTGGAGCGAGTCTT 101
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
46 erArgThrCysGlyGlyAlaSerTyrSerLeuArgArgCysLeuSer 62
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
102 CCAGAACCTGTGTGGGGTGCCTCTTATTCCTTGAGACGCTGCCTGAGC 151
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
63 SerLysSerCysGluGlyArgAsnIleArgTyrArgThrCysSerAsnVa 79
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
152 AGCAAGAGCTGTGAAGGAGAAATATTCGATATAGACATGCAGTATGT 201
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
79 lAspCysProGluAlaGlyAspPheArgAlaGlnGlnCysSerAlaH 96
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
202 GGACTGCCACCACGAGCGGTGATTTCCGAGCTCAACAGTGTCTGCTC 251
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
96 iAsnAspValLysHisGlyGlnPheTyrGluTrpLeuProValSer 112
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
252 ATAACGAGCTCAAGTACCATGGACAGCTTTATGAATGGCTTCCTGTATCT 301
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
113 AsnAspProAspAsnProCysSerLeuLysCysGlnAlaLysGlyThrTh 129
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
302 AATGACCCCGACATCATCTGCTCACTCAAGTCCCAACAAAGGAACAG 351
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
129 rLeuValValGluLeuAlaProLysValLeuAspGlyThrArgCysTyr 146
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
352 CCTGTTGTGGAAGTACGACCGAAGGCTTATAGATGGTACGGGATGCTATA 401
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
146 hrGluSerLeuAspMetCysIleSerGlyLeuCysGlnIleValGlyCys 162
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
402 CAGAACTCACTGGATATGTGCATCAGCGGCTATGCCAAATTTGTTGGCTGT 451
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
163 AspHisGlnLeuGlySerThrValLysGluAspAsnCysGlyValCysAs 179
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
452 CATCACAGCTGGGAGCAGCTGTCAAGAGGATAACTGTGGAGTGGCA 501
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
179 nGlyAspGlySerThrCysArgLeuValArgGlyGlnTyrLysSerGlnL 196
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
502 CGGAGATGGCTCAACGTCGCCGCTGGTCCGAGGCGCAGTATAAATCTCAGC 551
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
196 euSerAlaThrLysSerAspAspThrValValAlaIleProTyrGlySer 212
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
552 TCTCTGCAAGTAAATCGGATGACACCGTGGTGGCCATTCCTTACGGCAGC 601
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
213 ArgHisIleArgLeuValLeu 219
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
602 AGACACATCCGCCCTGTCTTA 622
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||

```

seq_name: gb_estl:A1917724

```

seq_documentation_block:
LOCUS      A1917724              568 bp      mRNA      linear      EST 14-DEC-1999
DEFINITION ttlic08.x1 NCI_CGAP_G66 Homo sapiens cDNA clone IMAGE:2240462 3'
            similar to WP:F25H8.3 CE05729 THROMBOSPONDIN LIKE ; mRNA sequence.
ACCESSION  A1917724
VERSION    A1917724.1  GI:5637579
KEYWORDS   EST.
SOURCE     human.
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE  1  (bases 1 to 568)
AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
JOURNAL   Unpublished (1997)
COMMENT   Contact: Robert Strausberg, Ph.D.
            Email: cgapb-1@mail.nih.gov
            Tissue Procurement: Christopher A. Moskaluk, M.D., Ph.D., Michael
            R. Emmert-Buck, M.D., Ph.D.
            cDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima
            Bonaldo, Ph.D.
            cDNA Library Arrayed by: Greg Lennon, Ph.D.

```

DNA Sequencing by: Washington University Genome Sequencing Center
 Clone distribution: NCI-CGAP clone distribution information can be
 found through the I.M.A.G.E. Consortium/LINL at:
www-bio.llnl.gov/bbrp/image/image.html

Insert Length: 922 Std Error: 0.00

Seq primer: -400P from Gibco

High quality sequence stop: 437.

FEATURES

source
 1..568
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:2240462"
 /clone_lib="NCI_CGAP_G66"
 /tissue_type="pooled germ cell tumors"
 /lab_host="DH10B"
 /notes="Vector: pT73D-Pac (Pharmacia) with a modified
 polylinker; Site.1: Not I; Site.2: Eco RI; Plasmid DNA
 from the normalized library NCI_CGAP_G64 was prepared, and
 ss circles were made in vitro. Following HAP purification,
 this DNA was used as tracer in a subtractive hybridization
 reaction. The driver was PCR-amplified cDNAs from a pool
 of 5,000 clones made from the same library (cloneIDs
 1257096-1258631, 1469064-1470983, and 1475592-1476743).
 Subtraction by Bento Soares and M. Fatima Bonaldo."

BASE COUNT 120 a 157 c 169 g 121 t 1 others
 ORIGIN

alignment_scores:

Quality: 1044.00 Length: 189
 Ratio: 5.583 Gaps: 0
 Percent Similarity: 98.942 Percent Identity: 98.413

alignment_block:

US-10-044-807-2 x A1917724/rev ..

Align seg 1/1 to reverse of: A1917724 from: 1 to: 568

```

666 CysProAlaArgTrpGluIleGlyLysTrpSerProCysSerLeuThrcy 682
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
568 TGCCCCAGCAAGTGGGANATTGGCAAGTGGAGTCCATGTAGTCTCACATG 519
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
682 sclyValGlyLeuGlnThrAspValPheCysSerHisLeuLeuSerA 699
|:::|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
518 TCGCGTTCGGCCTACAGACAGACAGCTCTCTGACGCCACCTGCTGTCCA 469
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
699 tGluMetAsnGluThrValIleLeuAlaAspGluLeuCysArgGlnPro 715
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
468 GAGAGATGAATGAACAGATCATCTGCTGATGAGCTGTGCGCCAGGCC 419
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
716 LysProSerThrValGlnAlaCysAsnArgPheAsnCysProProAlaTr 732
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
418 AAGCCAGCAGCGGTGCAAGCTTTGTAACCGCTTTAATTGCCGCCAGCCCTG 369
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
732 pTyrProAlaGlnTrpGlnProCysSerArgThrCysGlyGlyValG 749
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
368 GTACCCCTGCACAGTGCAGCCGTGTTCACAAACCTGCTGCGGGGGTTC 319
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
749 InLysArgGluValLeuCysLysGlnArgMetAlaAspGlySerPheLeu 765
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
318 AGAAACGTGAGGTCTTTTCCAAAGCAGCGCATGGCTGATGGCAGCTCCGTG 269
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
766 GluLeuProGluThrPheCysSerAlaSerLysProAlaCysGlnGlnAl 782
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
268 GAGCTCTCTGAGACCTCTGTTTCAGCTTCAAAACCTGCTGCCAGCAAGC 219
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
782 aCysLysLysAspAspCysProSerGluTrpLeuLeuSerAspTrpThrG 799
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
218 ATGCAAGAAAGATGACTGTCCAGCGAGTGGCTTCTCTCAGACTGGACAG 169
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
799 luCysSerThrSerCysGlyGluGlyThrGlnThrArgSerAlaIleCys 815
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
168 AGTGTTCACAAAGCTGCGGGGAAGGCCAGACTCGAAGCGCATTTGC 119
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||

```



```

34  lYLeuTrpAspAlaTrpGlyProTrpSerGluCysSerArgThrCysGly 50
|||||
205 GCCTCTGGGATGCTGGGGCCCTGGAGCGAGTGTCCAGAACCTGTGGT 254
51  GlyGlyAlaSerTyrSerLeuArgArgCysLeuSerSerLyssSerCysG1 67
|||||
255 GGGGGTGCCCTCTATTCCTTGAGACGCTGCTGAGCAGCAGCAGAGCTGTGA 304
67  uGlyArgAsnIleArgTyrArgThrCysSerAsnValAspCysProProG 84
|||||
305 AGGGAGAAATATTCGATATAGAACATGCAGTAATGTGGACTGCCACCAG 354
84  luAlaGlyAspPheArgAlaGlnGlnCysSerAlaHisAsnAspVallys 100
|||||
355 AAGCAGGTGATTCGAGCTCAACAGTGTCTGCTCATACACGACGTCAAG 404
101 HisHisGlyGlnPheTyrGlnTrpLeuProValSerAsnAspProAspAs 117
|||||
405 TACCATTGGACAGCTTATGTAATGGCTTCTGTATCTAATGACCCGACAA 454
117 nProCysSerLeuLysCysGlnAlaLysGlyThrThrLeuValValGluL 134
|||||
455 TCCATGCTCACTCAAGTGCACAAAGAAAGAACACGACCTGTTGTGGAAC 504
134 euAlaProLysValLeuAspGlyThrArgCysTyrThrGluSerLeuAsp 150
|||||
505 TAGCACCAGAGCTCTAGATGTGACGCGATGCTATACAGAATCACTGGAT 554
151 MetCysIleSerGlyLeuCysGlnIleValGlyCysAspHisGlnLeuG1 167
|||||
555 ATGTGCATCAGGGGCTATGCCAATTTGTTGCTGTGATCACCAGCTGGG 604
167 ySerThrVallysGluAspAsnCysGlyValCysAsnGlyAspGlySerT 184
|||||
605 GAGCACTGTCAAAAGAGATAACTGTGGAATGTGCAACGAGATGGCTCAA 654
184 hrCys 185
|||||
655 CGTC 659

```

seq_name: gb_est2:BI523618

seq_documentation_block:
LOCUS BI523618 581 bp mRNA linear EST 29-AUG-2001
DEFINITION 603051757F1 NIH_MGC_122 Homo sapiens cDNA clone IMAGE:5201199 5',
mRNA sequence.

ACCESSION BI523618

VERSION BI523618.1 GI:15348410

KEYWORDS EST.

SOURCE human.

ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

AUTHORS NIH-MGC

TITLE NIH-MGC

JOURNAL National Institutes of Health, Mammalian Gene Collection (MGC)

COMMENT Unpublished (1999)

Contact: Robert Strausberg, Ph.D.

Email: cgabbs-remail.nih.gov

Tissue Procurement: Life Technologies, Inc.

cDNA Library Preparation: Life Technologies, Inc.

cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)

DNA Sequencing by: Incyte Genomics, Inc.

Clone distribution: MGC clone distribution information can be

found through the I.M.A.G.E. Consortium/LLNL at:

http://image.llnl.gov

Plate: LLAM11504 row: b column: 16

High quality sequence stop: 577.

Location/Qualifiers

1..581

FEATURES

source

/organism="Homo sapiens"

/db_xref="taxon:9606"

/clone="IMAGE:5201199"

seq_name: gb_est1:AI342006

seq_documentation_block:

/clone_lib="NIH_MGC_122"

/lab_host="DH10B"

/note="Organ: pooled lung and spleen; Vector: pCMV-SPORT6;
Site_1: NotI; Site_2: EcoRV (destroyed); RNA source
anonymous pool of 24 week female lung, 16 week female
spleen, and 20-22 week male spleens. Library is oligo-dT
primed and directionally cloned (EcoRV site is destroyed
upon cloning). Average insert size 1.4 kb, insert size
range 1-3 kb. Library is normalized and enriched for
full-length clones and was constructed by C. Gruber
(invitrogen). Research Genetics tracking code 026.. Note:
this is a NIH_MGC Library."

BASE COUNT 140 a 151 c 159 g 131 t
ORIGIN

alignment_scores:

Quality: 958.00 Length: 172

Ratio: 5.570 Gaps: 0

Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:

US-10-044-807-2 x BI523618

Align seg 1/1 to: BI523618 from: 1 to: 581

```

1  MetGluCysArgArgAlaThrProGlyThrLeuLeuPheLeuAl 17
|||||
66  ATGGAATGCTGCGTCGGCAACTCTGGCACACTGCTCTTCTTGGC 115
17  aPheLeuLeuLeuSerSerArgThrAlaArgSerGluGluAspArgAspG 34
|||||
116 TTCTCTGCTCTGAGTTCAGGACCCACGCTCCGAGGAGGACCCGGAGC 165
34  lYLeuTrpAspAlaTrpGlyProTrpSerGluCysSerArgThrCysGly 50
|||||
166 GCCTATGGATGCTGGGGCCCATGGAGTGAATGCTCAGCACCTCGGG 215
51  GlyGlyAlaSerTyrSerLeuArgArgCysLeuSerSerLyssSerCysG1 67
|||||
216 GGAGGGGCTCTCTACTCTGTAGGGCTGCTGAGCAGCAAGAGCTGTCA 265
67  uGlyArgAsnIleArgTyrArgThrCysSerAsnValAspCysProProG 84
|||||
266 AGGAGAATAATCCCATACAGAACATGCAGTAATGTGGACTGCCACCAG 315
84  luAlaGlyAspPheArgAlaGlnGlnCysSerAlaHisAsnAspVallys 100
|||||
316 AAGCAGGTGATTCGAGCTCAGCAATGCTCAGCTCATATGATGTCAG 365
101 HisHisGlyGlnPheTyrGlnTrpLeuProValSerAsnAspProAspAs 117
|||||
366 CACCATGGCCAGTTTATGAATGGCTTCTGTGTCTAATGACCCCTGACAA 415
117 nProCysSerLeuLysCysGlnAlaLysGlyThrThrLeuValValGluL 134
|||||
416 CCCATGTTCACTCAAGTCCCAAGCCAAAGGAACAACCTGGTTGTGAAC 465
134 euAlaProLysValLeuAspGlyThrArgCysTyrThrGluSerLeuAsp 150
|||||
466 TAGCACCTTAAGTCTTAGATGGTAGCGCTTGTATACAGAACTTTGGAT 515
151 MetCysIleSerGlyLeuCysGlnIleValGlyCysAspHisGlnLeuG1 167
|||||
516 ATGTGCATCAGTGGTTATGCCAAATTTGTGGCTCCGATCACCAGCTGG 565
167 ySerThrVallysGlu 172
|||||
566 AAGCACCGTCAAGGAA 581

```

```

LOCUS      AI342006      522 bp      mRNA      linear      EST 29-DEC-1998
DEFINITION  qs98d10.x1 NCI-CGAP_G4 Homo sapiens cDNA clone IMAGE:1946131 3'
            similar to WP:F25H8.3 CE05729 THROMBOSPONDIN LIKE 1, mRNA sequence.
ACCESSION  AI342006
VERSION    AI342006.1 GI:4078933
KEYWORDS   EST.
SOURCE     human.
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE  1 (bases 1 to 522)
            NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
            National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
            Unpublished (1997)
COMMENT    Contact: Robert Strausberg, Ph.D.
            Email: cqabpsr@mail.nih.gov
            Tissue Procurement: Christopher A. Moskaluk, M.D., Ph.D., Michael
            Emmert-Buck, M.D., Ph.D.
            cDNA Library Preparation: M. Bento Soares, Ph.D.
            cDNA Library Arrayed by: Greg Lennon, Ph.D.
            DNA Sequencing by: Washington University Genome Sequencing Center
            Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            www-bio.llnl.gov/bbrp/image/image.html
            Seq primer: -400P from Gibco
            High quality sequence stop: 482.
FEATURES             Location/Qualifiers
     source           1..522
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:1946131"
                     /clone_lib="NCI-CGAP_G4"
                     /tissue_type="pooled germ cell tumors"
                     /lab_host="DH108"
                     /note="Vector: p773D-Pac (Pharmacia) with a modified
                     polylinker; 1st strand cDNA was prepared from 3 pooled
                     germ cell tumors, and was then primed with a Not I -
                     oligo(dT) primer. Double-stranded cDNA was ligated to Eco
                     RI adaptors (Pharmacia), digested with Not I and cloned
                     into the Not I and Eco RI sites of the modified p773
                     vector. Library is normalized. Library was constructed by
                     Bento Soares and M. Fatima Bonaldo."
     BASE COUNT      106 a 139 c 166 g 110 t 1 others
     ORIGIN
alignment_scores:
    Quality: 956.00      Length: 173
    Ratio: 5.558         Gaps: 0
    Percent Similarity: 99.422      Percent Identity: 99.422
alignment_block:
US-10-044-807-2 x AI342006/rev ..
Align seg 1/1 to reverse of: AI342006 from: 1 to: 522
693 CysSerHisLeuLeuSerArgGluMetAsnGluThrValIleLeuAlaAs 709
|||||
521 TCGAGCCACCTGCTTCCAGAGAGATGAATGAACAGTCATCCTGGCTGA 472
|||||
709 pGluLeuCysArgGlnPnProLysProSerThrValGlnAlaCysAsnArgp 726
|||||
471 TGAGCTGTGTGCGCCAGCNCNCAAGCCAGCAGCGGTGCAAGCTTGTAAACCGCT 422
|||||
726 heAsnCysProProAlaTrpTyrProAlaGlnPrpGlnProCysSerArg 742
|||||
421 TTAATTGCCCCCAGCTGGTACCCGTCACAGTGGCGCGCGTTCACGA 372
|||||
743 ThrCysGlyGlyValGlnLysArgGluValLeuLeuCysLysGlnArgMe 759
|||||
371 ACCTGTGGCGGGGTGTTTCAGAAACGTGAGGTCTTTTGAAGCAGCGCAT 322

```

```

759 tAlaAspGlySerPheLeuGluLeuProGluThrPheCysSerAlaSerL 776
|||||
321 GGCTGATGGCAGCTTCCCTGGAGCTTCTTGAGACCTTCGTTCAGCTTCAA 272
|||||
776 ysProAlaCysGlnAlaCysLysLysAspCysProSerGluTrp 792
|||||
271 AACCTGCTGCCAGCAGCATGCAAGAAAGATGACTGTCCAGCGAGTGG 222
|||||
793 LeuLeuSerAspTrpThrGluCysSerThrSerCysGlyGluGlyThrGl 809
|||||
221 CTCTCTCAGACTGGACAGAGTGTTCACAAAGCTCGGGGGAAGGCCCA 172
|||||
809 nThrArgSerAlaIleCysArgLysMetLeuLysThrGlyLeuSerThrV 826
|||||
171 GACTCGAAGCGCAATTCGCCAAGATGCTGAAAACCGCCCTCTCAACGG 122
|||||
826 alValAsnSerThrLeuCysProProLeuProPheSerSerSerIleArg 842
|||||
121 TTGTCAATTCCACCCCTGTGCCGCGCCCTGCTTCTTCTTCTCCATCAGG 72
|||||
843 ProCysMetLeuAlaThrCysAlaArgProGlyArgProSerThrLysHi 859
|||||
71 CCTGTATGCTGGCAACCTGTGCAAGCGCGCGCGGCCATCCACGAAGCA 22
|||||
859 sSerProHisIleAlaAla 865
|||||
21 CAGCCCGCACATCGCGGCC 3
seq_name: gb_estl:BB643318
seq_documentation_block:
LOCUS      BB643318      700 bp      mRNA      linear      EST 26-OCT-2001
DEFINITION  BB643318 RIKEN full-length enriched, 9.5 days embryo parthenogenote
            Mus musculus cDNA clone B130031C01 5', mRNA sequence.
ACCESSION  BB643318
VERSION    BB643318.1 GI:16478052
KEYWORDS   EST.
SOURCE     house mouse.
ORGANISM   Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
            1 (bases 1 to 700)
REFERENCE  1
AUTHORS    Arakawa,T., Carninci,P., Fukuda,S., Furuno,M., Hanagaki,T., Hara,A.,
            Hiramoto,K., Hori,F., Ishii,Y., Ito,M., Kawai,J., Konno,H., Kouda
            ,M., Koya,S., Matsuyama,T., Miyazaki,A., Nomura,K., Ohno,M., Sasaki
            Okazaki,Y., Okido,T., Saito,R., Sakai,C., Sakai,K., Sano,H., Sasaki
            ,D., Shibata,K., Shinagawa,A., Shiraki,T., Sogabe,Y., Suzuki,H.,
            Tagami,M., Tagawa,A., Takahashi,F., Takeda,Y., Tanaka,T., Toya,T.,
            Muramatsu,M. and Hayashizaki,Y.
            RIKEN Mouse ESTs (Arakawa,T., et al. 2001)
            Unpublished (2001)
            Contact: Yoshihide Hayashizaki
            Laboratory for Genome Exploration Research Group, RIKEN Genomic
            Sciences Center(GSC), Yokohama Institute
            The Institute of Physical and Chemical Research (RIKEN)
            1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
            Tel: 81-45-503-9222
            Fax: 81-45-503-9216
            Email: genome-res@gsc.riken.go.jp,
            URL: http://genome.gsc.riken.go.jp/,
            Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K., Itoh
            ,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
            Normalization and subtraction of cap-trapper-selected cDNAs to
            prepare full-length cDNA libraries for rapid discovery of new
            genes. Genome Res. 10 (10), 1617-1630 (2000)
            waki,K., Fujiwaka,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E.,
            Watahiki,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsuura
            ,S., Kawai,J., Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and
            Hayashizaki,Y.
            RIKEN integrated sequence analysis (RISA) system--384-format
            sequencing pipeline with 384 multicapillary sequencer. Genome Res.
            10 (11), 1757-1771 (2000)
            Konno,H., Fukunishi,Y., Shibata,K., Itoh,M., Carninci,P., Sugahara

```


'Y. and Hayashizaki, Y.
Computer-based methods for the mouse full-length cDNA
encyclopedia: real-time sequence clustering for construction of a
nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)
Kondo, S., Shinagawa, A., Saito, T., Kiyosawa, H., Yamanaka, I., Aizawa
, K., Fukuda, S., Hara, A., Itoh, M., Kawai, J., Shibata, K. and
Hayashizaki, Y.
Computational Analysis of Full-Length Mouse cDNAs Compared with
Human Genome Sequences. Mamm. Genome. 12, 673-677 (2001)
Please visit our web site (<http://genome.gsc.riken.go.jp>) for
further details.
e mouse tissues.

FEATURES
source
1..700
/organism="Mus musculus"
/db_xref="taxon:10090"
/clone_lib="B130031C01"
parthenogenote"
parthenogenote"
/tissue_type="parthenogenote"
/dev_stage="9.5 days embryo"
/lab_host="DH10B"
/note="Site_1: SalI; Site_2: BamHI; cDNA library was
prepared and sequenced in Mouse Genome Encyclopedia
Project of Genome Exploration Research Group in Riken
Genomic Sciences Center and Genome Science Laboratory in
RIKEN, Division of Experimental Animal Research in Riken
contributed to prepare mouse tissues. 1st strand cDNA was
primed with a primer [5'
GAGAGAGAGAGATCCAGAGCTCTTTTCTTTTCTTNN 3'], cDNA was
prepared by using trehalose thermo-activated reverse
transcriptase and subsequently enriched for full-length by
cap-trapper. cDNA went through one round of subtraction to
Ret = 229.0 second strand cDNA was prepared with the
primer adapter of sequence [5'
GAGAGAGAGATTCGAGTTAAATTAATCCCGCCCGCC 3'], cDNA
was cleaved with XhoI and BamHI. Vector: a modified
pBluescript KS(+) after bulk excision from Lambda FLIC I."
BASE COUNT 171 a 174 c 197 g 158 t
ORIGIN

alignment_scores:
Quality: 942.00 Length: 183
Ratio: 5.322 Gaps: 0
Percent Similarity: 96.721 Percent Identity: 92.896
alignment_block:
US-10-044-807-2 x BB643318 ..
Align seq 1/1 to: BB643318 from: 1 to: 700
1 MetGluCysArgArgAlaThrProGlyThrLeuLeuLeuPheLeuAl 17
|||||
40 ATGGAATGCTGCGGCGAGCTGCGCACACCGCTGCTCGTTCGCG 89
17 aPheLeuLeuLeuSerSerArgThrAlaArgSerGluGluAspArgAsp 34
90 TTTCCTACTCTGAGTTCAGGACTGCACGCTCAGAGAGACCGTGAAG 139
34 lyLeuTrpAspAlaTrpGlyProTrpSerGluCysSerArgThrCysGly 50
140 GCCTCTGGGATGCTGGGCGCTTGGGCGAGTGTTCCAGAACCTGTGGT 189
51 GlyGlyAlaSerTyrSerLeuArgCysLeuSerSerLyssSerCysG1 67
190 GGGGGTGCCTCTCTATTCCTGAGACGCTGCTCAGCAGCAAGAGCTGA 239
67 uGlyArgAsnIleArgTyrArgThrCysSerAsnValAspCysProG 84
240 AGGAGAGATTCGATATAGAACATGCAGTATGTGGACTGCCACCAG 289
84 luAlaGlyAspPheArgAlaGlnGlnCysSerAlaHisAsnAspValLy 100

|||||
290 AAGCAGGATGATTTCCAGCTCAACAGTTCCTGCTCATACAGCGTCAAG 339
101 HisHisGlyGlnPheTyrGluTrpLeuProValSerAsnAspProAspAs 117
:::|||||
340 TACCATGGACAGCTTTATGAATGGCTTCCTGTATCTATTAATGACCCGACAA 389
117 nProCysSerLeuLyScysGlnAlaLyScysGlyThrLeuValValGluL 134
TTCATGCTCAGTCAAGTCCAGCAAAAGGACCGCGGCTGTGGAAC 439
134 euAlaProLyScysValLeuAspGlyThrArgCysTyrThrGluSerLeuAsp 150
440 TAGCACCGAAGGCTTAGATGTCGCGATGCTATACAGAACTACCTGGAT 489
151 MetCysIleSerGlyLeuCysGlnIleValGlyCysAspHisGlnLeuG1 167
490 ATGTGTCATCAGCGGCTATGCCAAATTTGTCGTGATCACCCACTGGG 539
167 ySerThrValLyScysGlnAlaLyScysGlyValCysAsnGlyAspGlySer 183
540 GAGCACTGTCAAGAGGATACTGGGAGTGGGCAACGGAAATGGCTCA 588
seq_name: gb_est1:BB478282
seq_documentation_block:
LOCUS BB478282 619 bp mRNA linear EST 25-OCT-2001
DEFINITION BB478282 RIKEN full-length enriched, 13 days embryo heart Mus
musculus cDNA clone D330029102 3', mRNA sequence.
ACCSSION BB478282
VERSION BB478282 GI:16440420
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 619)
AUTHORS Arakawa, T., Carninci, P., Fukuda, S., Furuno, M., Hanagaki, T., Hara, A.,
, M., Koya, S., Matsuyama, T., Miyazaki, A., Nomura, K., Ohno, M.,
Okazaki, Y., Okido, T., Saito, R., Sakai, C., Sakai, K., Sano, H., Sasaki
, D., Shibata, K., Shinagawa, A., Shiraki, T., Sogabe, Y., Suzuki, H.,
Tagami, M., Tagawa, A., Takahashi, F., Takeda, Y., Tanaka, T., Toya, T.,
Muramatsu, M. and Hayashizaki, Y.
RIKEN Mouse ESTs (Arakawa, T., et al. 2001)
Unpublished (2001)
On Jul 23, 2000 this sequence version replaced gi:9395891.
Contact: Yoshihide Hayashizaki
Laboratory for Genome Exploration Research Group, RIKEN Genomic
Sciences Center (GSC), Yokohama Institute
The Institute of Physical and Chemical Research (RIKEN)
1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
Tel: 81-45-503-9222
Fax: 81-45-503-9216
Email: genome-res@gsc.riken.go.jp,
URL: <http://genome.gsc.riken.go.jp/>
Carninci, P., Shibata, Y., Hayatsu, M., Sugahara, Y., Shibata, K., Itoh
, M., Konno, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.
Normalization and subtraction of cap-trapper-selected cDNAs to
prepare full-length cDNA libraries for rapid discovery of new
genes. Genome Res. 10 (10), 1617-1630 (2000)
wagi, K., Fujiwara, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E.,
Watabiki, M., Yoneda, Y., Ishikawa, T., Ozawa, T., Tanaka, T., Matsura
, S., Kawai, J., Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and
Hayashizaki, Y.
RIKEN integrated sequence analysis (RISA) system--384-format
sequencing pipeline with 384 multicapillary sequencer. Genome Res. 10
(11), 1757-1771 (2000)
Konno, H., Fukunishi, Y., Shibata, K., Itoh, M., Carninci, P., Sugahara
, Y. and Hayashizaki, Y.
Computer-based methods for the mouse full-length cDNA
encyclopedia: real-time sequence clustering for construction of a
nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)

Ishii, Y. and Hayashizaki, Y.
Mapping of 19032 mouse cDNAs on mouse chromosomes. J. Struct.
Funct. Genomics 2 pre, I72-I86 (2001)
Please visit our web site (<http://genome.gsc.riken.go.jp>) for
further details.

FEATURES
SOURCE

```

FEATURES
source
e mouse tissues.
Location/Qualifiers
1. .633
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="5930437A14"
/clone_lib="RIKEN full-length enriched, 13 days embryo
forelimb"
/sex="mixed"
/tissue_type="forelimb"
/dev_stage="13 days embryo"
/lab_host="DH10B"
/note="Site_1: Sali; Site_2: BamHI; cDNA library was
prepared and sequenced in Mouse Genome Encyclopedia
Project of Genome Exploration Research Group in Riken
Genomic Sciences Center and Genome Science Laboratory in
RIKEN. Division of Experimental Animal Research in Riken
contributed to prepare mouse tissues. 1st strand cDNA was
primed with a primer [5'
GAGACAGAGAGATCCAGAGCTCTTTTCTTTTCTTTTCTTTTCTTTN 3'], cDNA was
prepared by using trehalose thermo-activated reverse
transcriptase and subsequently enriched for full-length by
cap-trapper. cDNA went through one round of normalization
to Rot = 10.0 and subtraction to Rot = 100.0. Second
strand cDNA was prepared with the primer adapter of
sequence [5' GAGACAGAGATCTCGAGTTAATTAATTCCTCCCCCCCCC
3']. cDNA was cloned into the XhoI and BamHI sites.
Vector: a modified pBluescript KS(+) after bulk excision
from Lambda FLC I. Cloning sites, 5' end: Sali; 3' end:

```

```
alignment_scores:
  Quality: 932.50
  Ratio: 4.639
  Percent Similarity: 95.261
  Gaps: 1
  Percent Identity: 87.204
  Length: 211
```

alignment block:

US-10-044-807-2 x BB620571 ..
Align seq 1/1 to: BB620571 from: 1 to: 633

1104 [euA]aG[uT]epheAroSerHicIouC[uw]icClnaCebHr ou r 1130

1104 LeuAlaGlnGluIlePheArgSerHisLeuCluHisGlnAspThrLeuLe
|||||:::|||||:::|||||:::|||||:::|||||
3 CTGGCCCAAGACATCTTCGAAGCCACCCTGGAGAACCGAGATCTGCTTCC 52

1120 uLysProSerGluArgThrSerProValThrLeuSerProHisLysH 1137

53 TAAGCCCTCTGAGCAGAGGTTCTCTCCCATGGCCGTCCTCCAGTCACAAAC 102

113/ isValSerGlyPheSerSerLeuArgThrSerSerThrGlyAspAla 1153
103 ATGTTCTCTGGCTTCACGACAGCTCCCAACG 1163

1154 GlyGlyGlySerArqArqProHisArqLysProThrIleLeuAqGLvsII 1170

150 GGAGTGGCTCCCGTAGGCCTCATCGAAGCCCGCCATCTTGCGGAAGAT 199

1170 eSerAlaAlaGlnGlnLeuSerAlaSerGluValValThrHisLeuGlyC 1187

200 CTCAGCGGCACAGCAACTCTCAGCCTCGGAGGTAGTCACCCACCTGGGGC 249

1187 lnThrValAlaLeuAlaSerGlyThrLeuSerValLeuLeuHisCysGlu
|||
250 |||

250 AGACIGIGGCCICTAGCCAGTGGGACCCCTGAGTGTGCTTCTGCACTGTGAG

1204 AlaAlGlyHisProArgProThrIleSerTTPalaAArgAsnGlyGlUg1 1220
GGGTAGGCCAACCCAAAGGCCACCACCTCCACTGGACCAGGAGAAGC 349

1220 uValGlnPheSerAspArqIlleLeuLeuGlnProaspSperLeuGlnI 1237
AGTTCAATTTCAGTGACAGAGATCTTCTCACGCCAGATGATTCCTTACAGA 399

1237 leLeuAlaProValGluAlaAspValGlyPheTyThrCysAsnAlaThr 1253
TCTTGGCACCAGCGAAGCTGATGTGGGTTCCTATCTTCAATGCACCA 449

1254 AsnAlaLeuGlyTyRasSerValSerlleAlaValThrLeuAlaGlyLy 1270
AATGCCTTGGGATATGACTCTCTCCCATGCTGCACATAGCAGAAA 499

1270 sProLeuVallYsThrSerAtqMetThrValIleAsnThrGluLysProA 1287
GCCACTAGTAAGACATCCCGAATGACGTGCTCAACACAGAGAAGCCTA 549

1287 laValThrValAspIleglySerThrIleLysThrValGlnGlyValAsn 1303
AAGTCCACAGTGGATATCGTGGCACCGTCAGAACCGTCCGAGGAGTGAAT 599

1304 ValThrIleAsnCysGlnValAlaGlyValPro 1314
GTGACAATTAACTGCCAAGTGGCAGGTGTACCT 632

seq_name: gb_est1:BB612331

seq_documentation_block:

LOCUS	BB612331		637 bp	mRNA	linear	EST 26-OCT-2001
DEFINITION	BB612331	RIKEN full-length enriched, 0 day neonate skin Mus musculus cDNA clone 4631431M22 5', mRNA sequence.				
ACCESSION	BB612331					
VERSION	BB612331.1	GI:16453331				
KEYWORDS	EST.					
SOURCE	house mouse.					
ORGANISM	Mus musculus					
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus. 1 (bases 1 to 637)					
AUTHORS	Arakawa,T., Carninci,P., Fukuda,S., Furuno,M., Hanagaki,T., Hara,A., Hiramoto,K., Hori,F., Ishii,Y., Ito,M., Kawai,J., Konno,H., Kouda,M., Koya,S., Matsuyama,T., Miyazaki,A., Nomura,K., Ohno,M., Okazaki,Y., Okido,T., Saito,R., Sakai,C., Sakai,K., Sano,H., Sasaki,D., Shibata,K., Shinagawa,A., Shiraki,I., Sogabe,Y., Suzuki,H., Tagami,M., Tagawa,A., Takahashi,F., Takeda,Y., Tanaka,T., Toya,T., Muramatsu,M. and Hayashizaki,Y. RIKEN Mouse ESTs (Arakawa,T., et al. 2001) Unpublished (2001)					
TITLE	Contact: Yoshihide Hayashizaki					
JOURNAL	Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center(GSC), Yokohama Institute					
COMMENT	The Institute of Physical and Chemical Research (RIKEN) 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan Tel: 81-45-503-9222 Fax: 81-45-503-9216 Email: genome-res@gsc.riken.go.jp, URL:http://genome.gsc.riken.go.jp/ Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K., Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y. Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes. Genome Res. 10 (10), 1617-1630 (2000) wagi,K., Fujiwake,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E., Wataniki,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsura,S., Kawai,J., Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and Hayashizaki,Y. RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multicapillary sequencer. Genome Res. 10 (11), 1757-1771 (2000)					

151 MetCysIleSerGlyLeuCys 157
|||||
530 ATGTGCAATCAGCGGCTATGC 550

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 24, 2002, 02:24:52 ; Search time 25.66 Seconds
(without alignments)
1677.237 Million cell updates/sec

Title: US-10-044-807-2
Perfect score: 9588
Sequence: 1 MECCRRATGCTLLFLAFL.....LKLCQLSQKRCRCGTCGKA 1762

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 231628 seqs, 24425594 residues

Total number of hits satisfying chosen parameters: 231628

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_AA:*
1: /cgn2_6/ptodata/2/iaa/5A_COMB.pep.*
2: /cgn2_6/ptodata/2/iaa/5B_COMB.pep.*
3: /cgn2_6/ptodata/2/iaa/6A_COMB.pep.*
4: /cgn2_6/ptodata/2/iaa/6B_COMB.pep.*
5: /cgn2_6/ptodata/2/iaa/PTUS_COMB.pep.*
6: /cgn2_6/ptodata/2/iaa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	373	3.9	788	2	US-08-918-914-4
2	359.5	3.7	441	3	US-08-985-526-3
3	311.5	3.2	1447	4	US-09-041-886-25
4	311.5	3.2	1447	5	PCT-US94-05277-2
5	307.5	3.2	1260	4	US-08-506-2968-21
6	304.5	3.2	469	1	US-08-313-2888-15
7	299	3.1	1253	4	US-08-506-2968-14
8	299	3.1	1268	4	US-08-506-2968-28
9	295	3.1	612	2	US-08-752-307B-11
10	294	3.0	1297	4	US-09-540-245A-17
11	290.5	3.0	615	2	US-08-752-307B-9
12	279.5	2.9	1651	4	US-09-540-245A-18
13	277.5	2.9	802	1	US-07-862-021B-12
14	277.5	2.9	802	1	US-07-862-021B-12
15	277.5	2.9	802	2	PCT-US93-03164-12
16	272.5	2.8	607	2	US-08-752-307B-12
17	264	2.8	807	1	US-07-862-021B-10
18	264	2.8	807	1	US-08-313-2888-10
19	264	2.8	807	5	PCT-US93-03164-10
20	261	2.7	869	1	US-08-374-834-16
21	261	2.7	869	2	US-08-644-271-29
22	257.5	2.7	630	2	US-08-752-307B-14
23	252.5	2.6	611	2	US-08-752-307B-10
24	247.5	2.6	868	1	US-08-374-834-1
25	247.5	2.6	868	2	US-08-644-271-1
26	246.5	2.6	596	2	US-08-752-307B-13
27	246	2.6	946	5	PCT-US95-08493-13

ALIGNMENTS

RESULT 1
US-08-918-914-4
; Sequence 4, Application US/08918914
; Patent No. 5876963
; GENERAL INFORMATION:
; APPLICANT: Mitchell, Peter
; APPLICANT: Hutchinson, Nancy
; APPLICANT: Lawton, Michael
; APPLICANT: Magna, Holly
; APPLICANT: Yocum, Sue
; APPLICANT: Murry, Lynn E.
; TITLE OF INVENTION: HUMAN NUCLEOTIDE PYROPHOSPHORYLASE
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESSES:
; ADDRESSEE: Incyte Pharmaceuticals, Inc.
; STREET: 3174 Porter Dr.
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/918,914
; FILING DATE: Filed Herewith
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Billings, Lucy J.
; REGISTRATION NUMBER: 36,749
; REFERENCE/DOCKET NUMBER: PF-0369
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-855-0555
; TELEFAX: 415-845-4166
; TELEX:
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 788 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; LIBRARY: GenBank
; CLONE: 1070094
US-08-918-914-4

Sequence 15, Appl
Sequence 19, Appl
Sequence 21, Appl
Sequence 15, Appl
Sequence 14, Appl
Sequence 14, Appl
Sequence 16, Appl
Sequence 20, Appl
Sequence 6, Appl
Sequence 6, Appl
Sequence 6, Appl
Sequence 2, Appl
Sequence 2, Appl
Sequence 2, Appl
Sequence 2, Appl
Sequence 3, Appl

28 245.5 2.6 478 5 PCT-US95-08493-15
29 245.5 2.6 860 5 PCT-US95-08493-19
30 245.5 2.6 868 5 PCT-US95-08493-21
31 240.5 2.5 1395 4 US-09-540-245A-15
32 240 2.5 568 1 US-07-862-021B-14
33 240 2.5 568 5 PCT-US93-03164-14
34 240 2.5 1381 4 US-09-540-245A-16
35 235.5 2.5 1170 1 US-08-313-2888-20
36 232 2.4 1018 1 US-08-408-093-6
37 232 2.4 1018 1 US-08-408-420A-6
38 232 2.4 1018 1 US-08-714-901-6
39 232 2.4 1018 3 US-08-040-741-6
40 231.5 2.4 917 1 US-08-245-295-2
41 231.5 2.4 917 1 US-08-481-130-2
42 231.5 2.4 917 1 US-08-656-984A-2
43 231.5 2.4 917 1 US-08-485-604-2
44 231.5 2.4 917 2 US-08-487-595-2
45 228.5 2.4 1501 2 US-08-447-464-3

TELECOMMUNICATION INFORMATION:
TELEPHONE: (302) 658-9141
TELEFAX: (302) 658-5613
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 441 amino acids
TYPE: amino acid
TOPOLOGY: linear
US-08-985-526-3

Query Match 3.7%; Score 359.5; DB 3; Length 441;
Best Local Similarity 25.6%; Pred. No. 1e-19;
Matches 138; Conservative 58; Mismatches 180; Indels 163; Gaps 29;

QY 324 CY----DLASNRVADQYC--HYYPENIKPKPKLQECNLDPC-----PASOGYQIMFYDL 373
Db 18 CYHNGVQYRNNEEWTDVSCTECHQNSVTICKVYSCNPATVPDGCPCRCWPSD- 76
QY 374 YHPLPWEA--TPWTACSSCGGQIQSRAYSCVEEDTQGHVTSVEEKKCMYTPKMPIAQ 431
Db 77 --SADDMGSPSEWSTSCSTSCGNIQORGRSCDSLNNRCGSSVQ-----TRT 122
QY 432 CNIFDCPKWLAQE-----WSPCTVTGGLRYRVVLCIDHRMHTG-----GCSPKTK 479
Db 123 CHIQECKRFRKQDGGWHSWSPSSCVTCGDGVTITRLCNPSPPQMNGKPCGEARETK 182
QY 480 PHKEECIVPTCPYKPEKLPVEAKL--PWFKQAQLEEGAASEPFIPEAWSACTVTC 538
Db 183 ACKKDAC-----PINGGWGPW-----SPWDICSVTC 208
QY 539 GVGTVQVIRVCOVLLSFQSVDLPIDCEGPKPASQ----RACYAGPCSGEIPFNPDE 594
Db 209 GGGVOKR-----SRICVDSRMTE--ENKELANELRRPLCY-----242
QY 595 TDGLFGGLQDFDELYEYEGFTKCSSECGGGVQEAUVVSCLNKOTREPAENLCVTSRRP 654
Db 243 ----HNGVQ--YENNEEWTVDSCTECH-----CQNSVTIC--KKVSCPIMP--CSNATVP 287
QY 655 ----POLLKSNLDPCPARWEIKWSPCSLTCGVGLQTRDVFCSHLLSREMNETVILAD 709
Db 288 DGECCPCWPSDSADDDGWSWP--SEWTSCTSCGNGIQORGRSCDSLNNRCGSSV----341
QY 710 ELCRQPKPSTVOACN--RENCPPAW--YPAQWOPCSTRTCGGVOKREVLCQK-----RMA 760
Db 342 ----QTRTCHIQECKRFRKQDGGWHSWSPSSCVTCGDGVTITRLCNPSPPQMNGKPC 397
QY 761 DGSFLELPETFCASAKPACQACKKDDCP-----SEWLLSDWTECSTSCGEGTQTRSAI 814
Db 398 EG---EARET-----KACKKDACPINGGWGPW--SPWDICSVTCGGVOKRSL 441

RESULT 3
US-09-041-886-25
; Sequence 25, Application us/09041886
; Patent No. 6235872
; GENERAL INFORMATION:
; APPLICANT: Bredesen, Dale E.
; APPLICANT: Rabizadeh, Sharoz
; TITLE OF INVENTION: Proapoptotic Peptides, Dependence
; TITLE OF INVENTION: Polypeptides and Methods of Use
; NUMBER OF SEQUENCES: 72
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Campbell & Flores LLP
; STREET: 4370 La Jolla Village Drive, Suite 700
; CITY: San Diego
; STATE: California
; COUNTRY: United States
; ZIP: 92122
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/985,526
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/608,845
; FILING DATE: 16-JUL-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: McMorrow Jr., Robert G

Query Match 3.9%; Score 373; DB 2; Length 788;
Best Local Similarity 25.3%; Pred. No. 2.3e-20;
Matches 133; Conservative 49; Mismatches 156; Indels 188; Gaps 29;

QY 341 YPENIKPKKLOECNLDPCPASDGYKQIMP-----YDLHYPLPWEATPTWATSSCGGGTQ 397
Db 429 YTRYRPAP-----PPPPACDGGQCVNPPVVGWHD---W--SDWSTCSTCGDGAK 476
QY 398 SRVAVCEVEDIOGHVTSVEEKKCMYTPKMPIAOPCNIFDCPKWLAQ--FWSPCTVTTCGGQL 456
Db 477 SRRRECGSTNNCGADYETE-----PCNLGPCQTWSEWCEWSTCSASCGSQ 522
QY 457 RYRVVLCIDHRMHTGCGSPKTKPHKEECIVPTCPYKPEKLPVEAKLPWFKAQAELEE 516
Db 523 RETRTEC-----HLGTNRCGKDYSEQCSA--GPC-----PENSOWED---559
QY 517 GAAVSEEPSFIPEAWSACTVTCGVGTQVIRVCOVLLSFQSVDLPIDB--CEGPKPASQ 575
Db 560 -----WQCSVTCGGVAVRQRTC-----LGGVFGDHLCCQPK--TEQ 595
QY 576 RACYAGPCSGEIPFNPDETDGLFGGLQDFDELYDWEYEGFTKCSKSGGGVQEAUVVSC- 634
Db 596 RACDGGPCSLWSB-----WQEWSTCSASCGSGMKRRQVQC 631
QY 635 LNKOTREPAENLCVTSRRPPQLKSNLDPCPARW--EIGKWSPCSITCGVGLQTRDVF 693
Db 632 FGTDCCQPNES-----QFCYGPCC--AEWTEWCEWSGCSKCGPGQRTTRGC 678
QY 694 SHLLSREMNETVILADELCRQPKSTV-----QACNRENCPPAWYPAQWOPCSTRTCGGVQ 749
Db 679 LGPNGQEA-----TCQGPSIETTLCEGQSCCNWS---EW--CHWSMCDKECGGG--723
QY 750 KREVLCQKQWADSFLELPETFCASAKPACQACKKDDCPSEWLLSDWTECST--SCGEG 807
Db 724 ----QVRYIEYMF-----RTGC-----EWSPCSTQLACEVG 750
QY 808 TOTRSALCRKMLTGLSTV-----VNSTLCPPPLPFFSSIRPC 844
Db 751 VQSRQOC--VGESECHCIGLAESQOCRLGTCQPKP-----PC 788

US-08-985-526-3
; Sequence 3, Application us/08985526
; Patent No. 6080728
; GENERAL INFORMATION:
; APPLICANT: Mixson, James A
; TITLE OF INVENTION: CARRIER:DNA COMPLEXES CONTAINING DNA
; TITLE OF INVENTION: ENCODING ANTI-ANGIOGENIC PEPTIDES AND THEIR USE IN GENE
; TITLE OF INVENTION: THERAPY
; NUMBER OF SEQUENCES: 43
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Connolly, Bove, Lodge, & Hutz
; STREET: 1220 Market Street, P.O. Box 2207
; CITY: Wilmington
; STATE: Delaware
; COUNTRY: U.S.A.
; ZIP: 19899
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/985,526
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/608,845
; FILING DATE: 16-JUL-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: McMorrow Jr., Robert G

```

; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/041,886
; FILING DATE:
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Campbell, Cathryn A.
; REGISTRATION NUMBER: 31,815
; REFERENCE/DOCKET NUMBER: P-LJ 2626
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (619) 535-9001
; TELEFAX: (619) 535-8949
; INFORMATION FOR SEQ ID NO: 25:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1447 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-09-041-886-25

```

```

Query Match          3.2%; Score 311.5; DB 4; Length 1447;
Best Local Similarity 20.7%; Pred. No. 3.9e-15;
Matches 151; Conservative 89; Mismatches 215; Indels 275; Gaps 32;

```

```

QY 1148 SSTGDAGGSRPHKPTILRKISAA---QQLSASEVVT-HLGTVALASGTLVLLHCE 1203
Db 119 ASLGDGSSIIISR-----TAKVAVAGPLRFLSQTESVTFAMGDT-----VLLKCE 162
QY 1204 AIGHPRPTISWANGEE---VQFSDRILLQDDSLQILAPVEADVGYTCNATNALGELTES 1260
Db 163 VIGEPMTIHWKQNDLTPIGDSRVVVLPSGALQISRLQPGDIGIYRCSARNP----- 217
QY 1261 VSTAVILACKPLVKTSRMVTINTEKPAVTVDIG-----STIKTVQGVNVTINLQC 1309
Db 218 ----ASSRTG-----NEAEVRILSDPGLRHQLYFLQRPNSVVAIEGKDAVLECC 262
QY 1310 VAGVPEAEVTFWFRNKSCL---GSPHLHEGS-LLLTNVSSDQGLYSCRAANLHGLTES 1365
Db 263 VSGYPPSPFTWLRGEVIOLRKSKYLLGGSNLLISNVTDDSGMTCVVYTKNENISAS 322
QY 1366 TOLLILDPPQVPTQLEDIRALLAATGPNLPSVLTSPGLTQLVDPGNSALLGCFIKGHPV 1425
Db 323 AELTVLVPPWF-----LNHPSNLYAYESMDIEFE-----CTVSGKPV 359
QY 1426 PNITWPHGQOPIVATGLTHHIIAAGIIOVANILOGSGOGFSCIAQNEAGVILMQKASLV 1485
Db 360 PTVMNMKNGDVVIPS---DYFQIVGGSNLRILGVKSDGEGFYQCVAEAGNAQTSQALI 416
QY 1486 IQDYVWSVDRLATCSASCGNRGVQOQPLRCLLNSTEVNFAHCAGKVRPA-----VQPIACN 1541
Db 417 VP-----KPAIPSSSVILPSAPRDVVPVIVS 441
QY 1542 RRDPSRMVMTWSACTRSCGGGVOT-----RRVTCOKLKAS 1578
Db 442 ----SRFVRLSWRP-PAEAKNGIQITFTVFSSREGDNRREALNTPQGSLSQITVGNLKPE 495
QY 1579 GLST-----PVSDNMCQTQAKRP-----VDTQACN---QQICVFEWAFSSWQ 1617
Db 496 AMYTFRVVAYNEWGGESSQPIKQVATQPELQVPGPVENLQAVSTSTSLITWEPPAY-- 553
QY 1618 CNGPCIGPHILAVOHRQVFCQ-----1637
Db 554 ANGP-----VQGRFLCTEVSSTKEQNIENVDGISYKLEGLKFKTEYSRLFLAYNRYGP 606
QY 1638 --TRDGIT-----LPSE-----QCSALPRPVSTONCWSEACSVHWRV-- 1672
Db 607 GYSTDDITWTLSDVPSAPPQNVSLVNVNSRIKYSWLPPLPPSGTONGFTGTGKIRHKTT 666
QY 1673 -----SLWTLCTA--TCGNYGFOSSRREVCVHARTKAVPEHLCSWGPBPANWOR 1719
Db 667 RRGEMETLEPNNLWTLFGLEKGSQYSFQ-----VSAMTVNGT-----GP-PSNWT 712

```

```

QY 1720 CNITPCENME 1729
Db 713 AE-TPENDLD 721

RESULT 4
PCT-US94-05277-2
; Sequence 2, Application PC/TUS9405277
; GENERAL INFORMATION:
; APPLICANT: Bruskin, Arthur
; APPLICANT: Jarosz, David E.
; APPLICANT: Johnson, Karen
; APPLICANT: Kinzler, Kenneth W.
; APPLICANT: Vogelstein, Bert
; APPLICANT: Zablrecky, James R.
; TITLE OF INVENTION: Antibodies Specific for DCC Gene Product
; NUMBER OF SEQUENCES: 2
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Banner, Birch, McKie & Beckett
; STREET: 1001 G Street, N.W.
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20001
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US94/05277
; FILING DATE:
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Kagan, Sarah A.
; REGISTRATION NUMBER: 32,141
; REFERENCE/DOCKET NUMBER: 01107.42709
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202.508.9100
; TELEFAX: 202.508.9299
; TELEX: 197430 HBMB UT
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1447 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; PCT-US94-05277-2

```

```

Query Match          3.2%; Score 311.5; DB 5; Length 1447;
Best Local Similarity 20.7%; Pred. No. 3.9e-15;
Matches 151; Conservative 89; Mismatches 215; Indels 275; Gaps 32;
QY 1148 SSTGDAGGSRPHKPTILRKISAA---QQLSASEVVT-HLGTVALASGTLVLLHCE 1203
Db 119 ASLGDGSSIIISR-----TAKVAVAGPLRFLSQTESVTFAMGDT-----VLLKCE 162
QY 1204 AIGHPRPTISWANGEE---VQFSDRILLQDDSLQILAPVEADVGYTCNATNALGELTES 1260
Db 163 VIGEPMTIHWKQNDLTPIGDSRVVVLPSGALQISRLQPGDIGIYRCSARNP----- 217
QY 1261 VSTAVILACKPLVKTSRMVTINTEKPAVTVDIG-----STIKTVQGVNVTINQ 1309
Db 218 ----ASSRTG-----NEAEVRILSDPGLRHQLYFLQRPNSVVAIEGKDAVLECC 262
QY 1310 VAGVPEAEVTFWFRNKSCL---GSPHLHEGS-LLLTNVSSDQGLYSCRAANLHGLTES 1365
Db 263 VSGYPPSPFTWLRGEVIOLRKSKYLLGGSNLLISNVTDDSGMTCVVYTKNENISAS 322
QY 1366 TOLLILDPPQVPTQLEDIRALLAATGPNLPSVLTSPGLTQLVDPGNSALLGCFIKGHPV 1425
Db 323 AELTVLVPPWF-----LNHPSNLYAYESMDIEFE-----CTVSGKPV 359

```


RESULT 6

US-08-313-288B-15
 ; Sequence 15, Application US/08313288B
 ; Patent No. 5750502
 ; GENERAL INFORMATION:
 ; APPLICANT: Jessell, Thomas M. and Avihu Klar
 ; TITLE OF INVENTION: CLONING, EXPRESSION AND USES OF A
 ; TITLE OF INVENTION: NOVEL SECRETED PROTEIN, F-SPONDIN
 ; NUMBER OF SEQUENCES: 20
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Cooper & Dunham LLP
 ; STREET: 1185 Avenue of the Americas
 ; CITY: New York
 ; STATE: New York
 ; COUNTRY: USA
 ; ZIP: 10036
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: PatentIn Release #1.0, Version #1.30
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/313,288B
 ; FILING DATE: January 5, 1995
 ; CLASSIFICATION: 435
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: White, John P. 28,678
 ; REGISTRATION NUMBER: 40028-A-PCT-US
 ; REFERENCE/DOCKET NUMBER: 40028-A-PCT-US
 ; TELEPHONE: (212) 278-0400
 ; TELEFAX: (212) 391-0526
 ; TELEX:
 ; INFORMATION FOR SEQ ID NO: 15:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 469 amino acids
 ; TYPE: amino acid
 ; STRANDEDNESS: single
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: peptide
 ; HYPOTHETICAL: NO
 ; ANTI-SENSE: NO
 ; US-08-313-288B-15

Query Match 3.2%; Score 304.5; DB 1; Length 469;
 Best Local Similarity 23.8%; Pred. No. 2.2e-15;
 Matches 131; Conservative 42; Mismatches 164; Indels 213; Gaps 33;
 QY 388 CSSCGGGIQRSAVSCVEEDIQHVTVEEWKMYTP-----KMPAQPCNIFDCPKW- 440
 Db 43 CKGLGGV-----SVEDC-CLNTAFAYOKRSGGLQPCRC-----SPRWS 81
 QY 441 LAQWSPCTVTCGGGLRYRVVLCIDHRMHTGGSPKTKPHKECIVPTPKPKKLP 500
 Db 82 LWSTWAPCSVTCSEGSOLRYRRCVGWNQCSGVAPCT----- 119
 QY 501 VEAKLPWFQAELEGAASVEPSFTPE-----AWSACTVTCGVGTQVRLVRCQVL 552
 Db 120 -----LEWQJQA-----CEDQQCCPENGSGWGWPEPCSVTCSKGTTRRAC--- 163
 QY 553 LSFSSQSVADLPIDCEGPKPA-----SORACVAGPCSGELPENPDETGLFGGLQ 603
 Db 164 -----NHPAPKCGGCHCPQQAQESACDTQQVC-----PTHGAWATWGP----- 201
 QY 604 DFDELWMEYEGFTKSESGGGVQEAASVCLNKQTR-----EPAEENLCVTSRPPQ 656
 Db 202 -----WTPCSASCHGGPHEP-----KETRSKCSAPE-----SOKPPG 235
 QY 657 LKASC-----NLDCPARWEIGKW-----SPCLTCGVGLQTRDVFCSHLLSREM 702
 Db 236 --KPCPLAYEQRRCTGLPPCPVAGGWGVPVSPCVTCGLGLQTMQRTCNHPVQH-- 291

QY 703 ETVILADELCROKPKFSTVQACN-RFNCP-----PAWYPAQWQPCSR-----TCG--GGV 748
 Db 292 -----GGPFCAGDATRT-HICNTAVPCPVQGEWDSW--GEWSPCIRRNMKISQCEIPGQ 343
 QY 749 QKREVLCQKRMADGSFLELPETFCASAKPACQACKKDDCP-----SEWLLSDWTEGSTS 803
 Db 344 QSRGRTCRGPKFDGH-----RCAGQQQDIRHCYSIQHCHPLKGSWSEW--STWGLCMPP 394
 QY 804 CGEG-TQTRSARICRMLKLTGLSTV-----VNSTLC-PPLPFSSSI-----RP 843
 Db 395 CGPNPTRAQRQLCTPLIPKYPPTVSMVGEQKKNVTFWGRPLPRCEELQCKLVVEKRP 454
 QY 844 CM-LATCARP 852
 Db 455 CLHVPACKDP 464
 RESULT 7
 US-08-506-296B-14
 ; Sequence 14, Application US/08506296B
 ; Patent No. 6313265
 ; GENERAL INFORMATION:
 ; APPLICANT: Phillips, Greg
 ; APPLICANT: Cunningham, Bruce A.
 ; APPLICANT: Crossin, Kathryn L.
 ; TITLE OF INVENTION: NEURITE OUTGROWTH-PROMOTING POLYPEPTIDES
 ; TITLE OF INVENTION: CONTAINING FIBRONECTIN TYPE III REPEATS AND METHODS OF USE
 ; NUMBER OF SEQUENCES: 77
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: The Scripps Research Institute
 ; STREET: 10550 No. 6313265th Torrey Pines Road, TPC-8
 ; CITY: La Jolla
 ; STATE: California
 ; COUNTRY: U.S.
 ; ZIP: 92037
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: PatentIn Release #1.0, Version #1.25
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/506,296B
 ; FILING DATE: 24-JUL-1995
 ; CLASSIFICATION: 514
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Fitting, Thomas
 ; REGISTRATION NUMBER: 34,163
 ; REFERENCE/DOCKET NUMBER: TSRI 488.0
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: (619) 554-2937
 ; TELEFAX: (619) 554-6312
 ; INFORMATION FOR SEQ ID NO: 14:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 1253 amino acids
 ; TYPE: amino acid
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: protein
 ; US-08-506-296B-14

Query Match 3.1%; Score 299; DB 4; Length 1253;
 Best Local Similarity 20.1%; Pred. No. 2.9e-14;
 Matches 187; Conservative 110; Mismatches 345; Indels 288; Gaps 36;

QY 843 PCMLATCARPGRSTKHSPIHAAARKVYIQTTRRORKLHFVVGGFAYLLPKTAVVLRCPAR 902
 Db 16 PCLLIQI-----PEEYEGHHVMEPPVITEQSPR-----LVVFPIDDISLKCEAS 60
 QY 903 RVKPLITTEKDOHLISSTHTVTVAPGYLKIHLKPSDAGVYTCSPAGAREHFVKLIG 962
 Db 61 GKPEVQFRWTRDGVHFKPKELGVTV-----QSPHSGSFTIUGN--NSNFAQRFQ 110
 QY 963 GNRKLVARPLSPREEVLAGRKGPKKEALQTHKHQNGIFSNKAEKRGKLAANPGSRD 1022

111 IYCFASNKLTAMSHIRLMBAGPK-----137
Db
1023 DLVSRLLQGGWPGELLASWEADQSNERTTSEEDGAEQV-----LLHLPTMTVTE 1074
QY
138 -----WPKETVKPVEVEGESVVLPCNPPPSAEPLRIYWMNSKILHIK-----QD 182
Db
1075 QRRLLDILGNLSQPELRDLYSKHLVAQAQIFRSHLEHODTLKPSERRTSPVTLSP 1134
QY
183 ERYTMGONGNLYFANVLTSDHSDY-----ICHNFPGTRTII-----QKEPI----- 225
Db
1135 HKHVSFSSSLRTSGDAGGRRRPHKPTILRKISAAQOLASAEVTHLQGVVALASG 1194
QY
226 -----DLRVKAT-----NSMIDRKPRLLFPTN-----SSSHLVALQGG----- 258
Db
1195 TLSVLLHCEAIGHPRPTISWARGEEVQFSDRILLO-PDSSLQTLAPVEADVGYTCNAT 1253
QY
259 --PIVLECIAGFPTPTIKWLPSGMP-ADRVTYQHNKTLQLLKYGEEDGGEYRCLAE 315
Db
1254 NALGYDSVSTAVTLAGPLVKTSRMVTINTEKPAVTVDIGSTIKTVQGVNNTINQVAGV 1313
QY
316 NSLGSARHAYVTEAPYW-----LHKP-----QSHLYGGETARLDCQVQGR 359
Db
1314 PEAETWFRNRSKLGSP-----HHLHGSLLLNTVSSSDQGLYSCRAANLHGETE 1364
QY
360 PQPVTWRIN---GIPVEELAKDQKRIQKALILSNVQPSDTMTVQCEARNRHGLLA 415
Db
1365 STQLLIDPP-----QVP-----TOLED----- 1382
QY
416 NAYIYVQLPAKILTDNQTYMAYOGSTAYILLKAPGAPVPSVQWIDEDGTVIQUERFF 475
Db
1383 -----IRAL-----LAATGNLPSVLTS-----PLGTQLVLDP-----GNSA 1414
QY
476 PYANGTIGIRLDANDGTREYCLAANDONNVTIMANLAKVDATQITQGRSTIEKKGSRV 535
Db
1415 ILGCPKIGHP--VPNTITWFGGQPIVATGLTHHILAAQILQVANLGGSGGEGESCLAQ 1472
QY
536 TFTQASDPISLQPSITWKGDRDLQELGDSKYPFIEDGRV-THSLDYSDQCNVSCVAS 594
Db
1473 NEAGVLMQKASLVIQDYWMSVDRLATCSACGNRGVQPRCLNLSNEVNPACHCAGVR 1532
QY
595 TELDVESRAQLLVVSGPVPRLVLSLH-----LITQSVRVSWSPADH 641
Db
1533 PAVQPIAC-----NRRDCPSRW-----MVTISWACTRSCGGVQ--TRRVTCKLKAAGI 1580
QY
642 NA--PIEKYDIEFEKEKAPEKWKYSLGKVPGNQJSTTLKLSYVHVTRVTAINKYGPGE 699
Db
1581 STPVSNDCMTOVA---KRPVDTQACNQ---LCVWEAFSSWQCQNGPCGTGPHLAYOHRQV 1634
QY
700 PSPVSETVVTPEAAPEKNDVDYKGCNETTNMTNMTWKPLRWDWNAPE-----QVQYRVQ 753
Db
1635 FCQTRDGITLPSEQSALPRPVSTQNCWSE 1664
QY
754 W-----RPOGTGPMQOE 765
Db
RESULT 8
US-08-506-296B-28
; Sequence 28, Application US/08506296B
; Patent No. 6313265
; GENERAL INFORMATION:
; APPLICANT: Phillips, Greg
; APPLICANT: Cunningham, Bruce A.
; APPLICANT: Crosslin, Kathryn L.
; APPLICANT: Crosslin, Kathryn L.
; TITLE OF INVENTION: NEURITE OUTGROWTH-PROMOTING POLYPEPTIDES
; NUMBER OF SEQUENCES: 77
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: The Scripps Research Institute
; STREET: 10550 No. 6313265th Torrey Pines Road, TPC-8
; CITY: La Jolla
; STATE: California
; COUNTRY: U.S.

ZIP: 92037
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/506,296B
; FILING DATE: 24-JUL-1995
; CLASSIFICATION: 514
; ATTORNEY/AGENT INFORMATION:
; NAME: Fitting, Thomas
; REGISTRATION NUMBER: 34,163
; REFERENCE/DOCKET NUMBER: TSRI 488.0
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (619) 554-2937
; TELEFAX: (619) 554-6312
; INFORMATION FOR SEQ ID NO: 28:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1268 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-506-296B-28
Query Match 3.1%; Score 299; DB 4; Length 1268;
Best Local Similarity 20.6%; Pred. No. 3e-14;
Matches 179; Conservative 119; Mismatches 327; Indels 246; Gaps 37;
QY 891 PXTAVVLPCPARRVRKPLITWEKDGQH--LISSTHTVAP-FGYLKIHLKPSDA---G 943
Db 55 PRENIVIOCEAKGKPPSPFSTWRTNGTHEDIDKDAQYMKPNSGTLVNVTMNGVKAAYEG 114
QY 944 VYTCAGAPAREHIVKILGNGKLVARP-LSPRSEEEVLAGRKGPGKEALQTHKHONGIF 1002
Db 115 VYQCTARNERGAII-----SNVIVRPSRL-----WTKKLEPNHVRG-- 155
QY 1003 SNGSKAEKRGLANPGSRVDDLVSRLEGGWPGGELLASWEADQSNERTTSEEDPGAQ 1062
Db 156 -----DSLVLNCRPPVGLPPTII-FW--MDNAFQL-----PQSER 188
QY 1063 VLLHLPTMTVQBORLDDILGNLSQPELRDLYSKHLVAQAQIFRSHLEHODTLKPK 1122
Db 189 V-----SQUNGDLIFSNOVQEDIRVDY-----ICYARFNHTQFI--- 223
QY 1123 SERRTSPVTLSPHKHVSFGFSSSLRTSSTGDAGGSRPHKPTILRKISAAQOLASAEV 1182
Db 224 --QKQKQPIIS-----VKVFSTKPTV-----ERPPVLL----- 247
QY 1183 THLGQT---VALASGTLVLLHCEAIGHPRPTISWARGEEVQFSDRILLOPDDSLQILA 1239
Db 248 TPMGSTSNKVELRGVNL--LLECIAAGLPTPVIRWIKEGEGELPANRTFFENFKTKLIID 305
QY 1240 PVEADVGYTCNATNALGYDSVSIATVLACKPLVKTSRMVTINTEKPAVTVDIGSTIKTV 1299
Db 306 VSEADSGNKTCTARNILGSTHHVIVSVTKAAPTWTAPRNLVLS----- 349
QY 1300 QGVNVTINQVAGVPEAEVTFWRNRSKLG-----SPHILHEG-SLLLTNWSSSDQGLYSR 1354
Db 350 PGEDGILLICRANGNPKPSISWLTNGVPIALAPEDPSKVDGDTIIFSAVQERSAVYQCN 409
QY 1355 AANLHGETESTOLLIL-DPPQVPTQLEDIRALLAAGPNLPSVLTSPGTQLVLDPGNS 1413
Db 410 ASNEYGYLLANAFVNLAEPRILTPANKLYQVIA-----DSP----- 447
QY 1414 ALLGCPKIGHPVNIWTFHGGQPIVATGLTHHILAAQILQVANLGGSGGEGESCLAQ 1473
Db 448 ALIDCAFYGPKPEIENFWRGVKGSILRG--NEYVFDHNGTLEIPVAKDSTGTGTVARN 505
QY 1474 EAGVLMQKASLVIQDYWMSVDRLATCSACGNRGVQPRCLNLSNEVNPACHCAGVR 1533
Db 506 KLGKTONQEVQLEVKD-----PTMTIKOPOYKVIQORSQAAS-FECVTKHDP 549

QY 1534 AVOPIAC-----NRRDCPSRWV-----TWSACTRSCGGVOTRVVTCQKLGASISTP 1583
DB 550 TLIPVWLNKDNELPDDREFVGGKDNLTIMNVTDKDDGTTCIVNTTLDSSASALTV 609
QY 1584 VSDMCTOV---AKRPVDTQACNQ--OLCVWAFSSWGCGPC1-----1623
DB 610 VAAPPTPAILIARPNPDLDELTLGQLERSIELSWPGGEENNSPITNFVIEYEDGLHEPGV 669
QY 1624 -----GPHLAVOHR-----OVFCQTRDGIITLPEQO-----SALP--RPVST 1658
DB 670 WHYQTEVPGSHTTVOLKLSPPYNYFRVIAVNEIGRSQSEFSEQLTKASAPDENPSNV 729
QY 1659 QNCWSEACSVHVRVSLWTLCTATCNGYGFQS 1689
DB 730 QGIGSEPDML---VITWESLK-----GFQS 751

RESULT 9
US-08-752-307B-11
; Sequence 11, Application US/08752307B
; Patent No. 5952171
; GENERAL INFORMATION:
; APPLICANT: McCarthy, Sean A.
; APPLICANT: Gearing, David P.
; APPLICANT: Levinson, Douglas A.
; TITLE OF INVENTION: METHOD FOR IDENTIFYING GENES
; TITLE OF INVENTION: ENCODING NOVEL SECRETED OR MEMBRANE-ASSOCIATED PROTEIN
; NUMBER OF SEQUENCES: 14
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson, P.C.
; STREET: 225 Franklin Street
; CITY: Boston
; STATE: MA
; COUNTRY: US
; ZIP: 02110-2804
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: Windows95
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/752,307B
; FILING DATE: 19-NOV-1996
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Meiklejohn, Ph.D., Anita L.
; REGISTRATION NUMBER: 35,283
; REFERENCE/DOCKET NUMBER: 09404/020001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617-542-8906
; TELEFAX: 617-542-5070
; TELEX: 200154
; INFORMATION FOR SEQ ID NO: 11:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 612 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-752-307B-11

Query Match 3.1%; Score 295; DB 2; Length 612;
Best Local Similarity 21.2%; Pred. No. 1.9e-14;
Matches 141; Conservative 98; Mismatches 243; Indels 182; Gaps 27;
QY 891 PXTAVVLRCPARRVRKPLTWKDKQOH--LISSHTVTVAP--FGYLKIHLRLKPSDA---G 943
DB 55 PRENIVIOCEAKGKPPPSFSWTRNGTHFDIDKDAQVTKPNSGTLVNMNGVKAFAVEG 114

QY 944 VYCSAGPAREHFVIKLIGNRKLYARP--LSPRSEEVLAGRGGPKALQTHKHQMGIF 1002
DB 115 VYCTANERGAAI-----SNNIVIRPSRSL-----WKEKLEPNHVRG-- 155
QY 1003 SNGSKAEKRGKLAANPGSRDYDDLVSRLLEQGGWPGELLASWFAODSARNRTISEDPGAEQ 1062
DB 156 -----DSVLNCRPPVGLPPPII--FW--MDNAFQRL-----PQSER 188
QY 1063 VLLHLFTVMTEORRLDDILGNLSQOPEELRDLYSKHLAQLAQAEIFRSLEHODTLKP 1122
DB 189 V-----SOGLNGDLYFSNVQPEDTRVDY-----ICYARFNQTII--- 223
QY 1123 SERRTSPVTLSPHKHVSFGFSSSLRTSTGDAGGSRPRHKKPILRKISAAQOLSASEV 1182
DB 224 ---CQKQPIIS-----VKVFSTKPVV-----ERPPVLL----- 247
QY 1183 THLGOT---VALASGTLVLLHCEAIGHPRPTISWARGNEEVQFSDRILLOPDDSLQILA 1239
DB 248 TPMGSTSNKVELRGNVL---LLECIAAGLTPVIRWKEGGELPANRTFFENFKTKLIIID 305
QY 1240 PVBADVGFTCNATNALGYDSVSIATVLAKPLVKTSRMVTINTEKPAVTVDIGSTIKTV 1299
DB 306 VSEADSGNYKCTARTLTGSTHHVSVTVKAAPYWITAPRNVLVS----- 349
QY 1300 QGVNVTINCOVAGVPEAEVTFWRNKSGLG-----SPHLLHEG--SLLLTNVSSSDOGLYSCR 1354
DB 350 PGEDTLCIRANGNPKPSISWTNGVPIAIPEDSRKVDGTIIIFSVAQVRESSAVYQCN 409
QY 1355 AANLHGELTESTQLLIL-DPPQVPTQLEDIRALLAATGPNLPSVLTSPLGTOLVLDGNS 1413
DB 410 ASNEYGYLLANAFVNVLAEPPIRLTPANKLQVIA-----DSP----- 447
QY 1414 ALLGCPKIGHVPNTWPHGQPIVITATGLTHILAAGQILQVANLSGSGCEFSCLAQN 1473
DB 448 ALIDCAVFGSPKPELENERGVKGSILRG--NEYVPHDNGTLEIPVAKDSTGTYTCVARN 505
QY 1474 EAGVLMOKASLVIODYWWSVDRLATCSASCNGRVQOQPRLCCLLNSTEVNPAHCAGKVRP 1533
DB 506 KLGTONEVQLEVKD-----PTMIKQPYKVIORSQAQAS--FECVIKHPD 549
QY 1534 AVQP 1537
DB 550 TLIP 553

RESULT 10
US-09-540-245A-17
; Sequence 17, Application US/09540245A
; Patent No. 6270984
; GENERAL INFORMATION:
; APPLICANT: Goodman, Corey
; APPLICANT: Kid, Thomas
; APPLICANT: Brosse, Katja
; TITLE OF INVENTION: Modulating Robo: Ligand Interactions
; FILE REFERENCE: B98-031-3
; CURRENT APPLICATION NUMBER: US/09/540,245A
; PRIOR FILING DATE: 2000-03-31
; PRIOR APPLICATION NUMBER: 60/065,544
; PRIOR FILING DATE: 1997-11-14
; PRIOR APPLICATION NUMBER: 60/081,057
; PRIOR FILING DATE: 1998-04-07
; NUMBER OF SEQ ID NOS: 20
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 17
; LENGTH: 1297
; TYPE: PRT
; ORGANISM: Caenorhabditis elegans
US-09-540-245A-17

Query Match 3.1%; Score 294; DB 4; Length 1297;
Best Local Similarity 22.0%; Pred. No. 7.6e-14;

Wed Jul 24 11:35:40 2002

```
Matches 142; Conservative 78; Mismatches 224; Indels 200; Gaps 26;
QY 897 LRCPARVRKP---LITWEKQGHILS-----STHTVAPFGYKLIHRL-----KPSDA 942
Db 49 LNCGA---KPSAKITWKDGQPVITNKEQVNSHRIVLDTGSLFLKYNKNGKSDSA 104
QY 943 GYVTSAGAREHFVTKLGGNKKLVAR---PLSPRSEEVLAGKGGPKALQPHKHQ 998
Db 105 GAYYCAV---SNEHGEVSKNEGSLKLAHREDFRVRPTVQAL-----GEMAVLECSPPR 157
QY 999 NGIFSNKSAEKRAAGLAANPGSYDDLVSLRLLEQGGWPGGELLASWEAQDSAEARTTSEEDP 1058
Db 158 -----IQDMPRYTLHSD-----GNLIIDPVDSDSGTYQCVAN----- 174
QY 1059 GADQVILLHP-FVMVTEQRLDILGNLSQOPEELDLYSKHLVAQLAQEIFRSHLEHQD 1117
Db 175 -----IQDMPRYTLHSD-----GNLIIDPVDSDSGTYQCVAN----- 207
QY 1118 TLKPKSPRRTSPVTLSPHKHHVSGFSSSLRTSSTGDAGGSGRRPRPKPTILRKISAAOOLS 1177
Db 208 --NMGERVSNPARLSVFE-----KPKF-----EQE 231
QY 1178 ASBVTHLGOTVALASGTLVSLHCEAIGHPRPTISWARNGEVEQFSDRILLQDDSLQI 1237
Db 232 PKDMTVDVG-----AAVDFCRVTGDPQQLTWKRNKNEPMPVTRAYIAKDNRLRI 282
QY 1238 LAPVEADVGYTCNATNALGYSVIAVTLAGKPLVKTSMVTINTEKPA-VTVDIGSTI 1296
Db 283 ERVQPSDEGEVGYARNPAGTLEASAHLRVQAPFSQT-----KPADQSPAGGT- 332
QY 1297 KTVQGVNVTINCOVAGVPEARVTFWR-NKSLGSPHLLH-----EGSLLLTNYSSS 1346
Db 333 -----ATECTLVGQSPAYFWSKEGQDQLLEPFSVSADGRUKVSPGTGLTIEEVROV 385
QY 1347 DQGLYSRAANLHGLTESTQLLLIDPPQVPTQLED-----TRALLATGP 1392
Db 386 DEGAYVCAGMNSAGSSLSKAALKATFETKGRVQKKSKMGKQKOKNVQSIILYKILISAVIG 445
QY 1393 NLPSVLSPLGT-----QVLDPCNSALLGCPKIGHVPVNTWPHGGOPT-VTATGLTH 1445
Db 446 NTPA---KPPPTIEHGHONQTLMGSSAILPCQASGKPTPGISWLRLDGLDITDSRISQ 502
QY 1446 HILAAQQLQVANLSSGSGQSFCSLAQNEAGVLMOKASLVIQDY 1489
Db 503 HSTGS---LHIALDKKPTGVYTCIAKNEDGESTWSASLITVEDH 543

RESULT 11
US-08-752-307B-9
; Sequence 9, Application US/08752307B
; Patent No. 5952171
; GENERAL INFORMATION:
; APPLICANT: McCarthy, Sean A.
; APPLICANT: Gearing, David P.
; APPLICANT: Levinson, Douglas A.
; TITLE OF INVENTION: METHOD FOR IDENTIFYING GENES
; TITLE OF INVENTION: ENCODING NOVEL SECRETED OR MEMBRANE-ASSOCIATED PROTEIN
; NUMBER OF SEQUENCES: 14
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson, P.C.
; STREET: 225 Franklin Street
; CITY: Boston
; STATE: MA
; COUNTRY: US
; ZIP: 02110-2804
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: Windows95
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/752,307B
; FILING DATE: 19-NOV-1996
```

```
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Meiklejohn, Ph.D., Anita L.
; REGISTRATION NUMBER: 35,283
; REFERENCE/DOCKET NUMBER: 09404/020001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617-542-5070
; TELEFAX: 617-542-8906
; TELEX: 200154
; INFORMATION FOR SEQ ID NO: 9:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 615 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-752-307B-9

Query Match 3.0%; Score 290.5; DB 2; Length 615;
Best Local Similarity 21.9%; Pred. No. 4.3e-14;
Matches 147; Conservative 88; Mismatches 242; Indels 195; Gaps 27;
QY 896 VLRCPARVRKPLITWEKDGQ-----HLISSTHTVAPFGYKLIHRLKPSDAGVVTCSAGP 951
Db 59 IIEEADGQPEPEYSWTKNGKKFDQAYDNRLKQPGRGTLVITPKDEDGRHYQCFA-- 116
QY 952 AREHFVILKIGNRKLVARPLSPRSEEEVLAKRGKGPKEALQTHKHONGIFSNKSAEKR 1011
Db 117 -----SNEGCTA-----TNSVYVVRKAELNAFKADEAAKTLE 147
QY 1012 GLAANPGSYDDLVSLRLLEQGGWPGGELLASWEAQDSAEARTTSEEDPQAEQVILLHPTM 1071
Db 148 AVEGEP-----FMLKCAAPDGFPSPTV-NWMIQESIDGSIKS----- 183
QY 1072 VTEQRRLDILGNLSQOPEELDLYSKHLVAQLAQEIFRSH-----LEHQDTLLKPS 1123
Db 184 INNSMTLDPGELNWFNSVNTREDASDDFYACSAISVFSEYKIGNKVLDDVKMGVSAS 243
QY 1124 ERTSPVTLSPHKHHVSGFSSSLRTSSTGDAG-----GSGRRRHRKPTILRKISAAQ 1174
Db 244 QNKHPV---ROYVSRROSALR-----GKRMELFCIYGTPLPQ---TWMSK--DQ 287
QY 1175 QLSASEVVT--HLGQT----- 1188
Db 288 RIQWSDRITQGHYKSLVIRQTNFDDAGTYTCDVSNVGNQAQSFSLILNVNVPYTKEP 347
QY 1189 -VALASGTLVLLHCEAIGHPRPTISWARNGEVEQFSD-----RILLQDDSLQILAPVEA 1243
Db 348 EIATAAEDEEVFECAAGVPEPKISWIINGKPIEQSTPNRPTV--TDNTRIINLVKG 405
QY 1244 DVGFYTCNATNALGYSVIAVTLAGKPLVKTSMVTINTEKPAVTVDIGSTIKTVQGVN 1303
Db 406 DTGNYCNATNSLGYVYKDYLVNVAEP-----PTI--SEAPA-----AVSTVDGRN 450
QY 1304 VTINQVAGVPEAEVTFWRNKSXL-GSPHLLH-EGSLLLTNYSSSDOGLYSCRAANLHGE 1361
Db 451 VTIKRVNGSPKPLVKWLRASNWLTTGGRYNVQANGLEIDODVTFSDAGKTYTQAKNFGE 510
QY 1362 LTESTOLLILDLP--POVPTOLEDIRALLAATGNPLPSVLTSPLCTQLVLDPGNSALLGCP 1419
Db 511 IQADGSLVWKEHTITQEPQNYE-----VAA-----GQSATFRCN 544
QY 1420 IKGHVPV---NITWPHGGOPTVATGLTHHILAAQGLILOVANLSSGSGQSFCSLAQNEAG 1476
Db 545 -EAHDDTLETEIDWKKDQSQI--DFEAQPFVKTNDNSLTIAKTMELDSGEYTCVARTRLD 602
QY 1477 VLMOKASLVIQD 1488
Db 603 EATARANLIVQD 614
```

QY	1020	RYDDLVSRLLEGGWPGELLAS-----WEAQDSAPRNTTSREDQAE	1061
Db	62	ROEDFPPIRIVEH---PSDLIVSKGEPATLNCABGRPTPTIEWKGERVETDKDDPRSH	118
QY	1062	QVIL---HLPTMTWTEORRL-----DDILGNLSQOPELRLDYSKH---	1099
Db	119	RMLLPSGSLFLRIVHGKRSRDPBGVYVCVARNYLGEAVSHNASIEUALLHDDFRQNPDS	178
QY	1100	-LVLAQAEIIF-----RSHLEHQDTLTK---PSERRTPSVTLSPHKHIVGFSSSLRTSST	1150
Db	179	VMVAEGEPANWECOPPHGHDEPTTISWKKDGSPLDKDKERIT-----IRGCKLMTITRK	232
QY	1151	GDAGG-----GSRPHRRKPTILRKISAAQOLASAEVWTHLGOTVALASGTLVLLHCPAIG	1206
Db	233	SDAGKYVCVGTNMGERESEVAELTVLERFSFKYRPSNLAVTV-----DDSABFKCEARG	287
QY	1207	HPRTISWARGCEVOFSRDLLOPDSOIGILAPVEADVGFTYCNATNALGYDSVSIAVT	1266
Db	288	DPVPTVMRRKDDGELPKS-RYEIRDDHTLAKIRVTAGDMGSYCTCAEPAVMYKAFASATIT	346
QY	1267	LAKGP--LVKTSRMTVINEKPAVTVDIGSTIKTVQGVNVTINCOVAGVPEAEVTFNRNK	1324
Db	347	VOEPPHEVVKPRQVV-----ALGRTVTFCEATGNPQPAIFNRRE--	387
QY	1325	SKLGSPPHL-----HBSGLLLTNVSSDGLYSCRAANLHGELTESTQLL	1369
Db	388	---GSONLLFSYQPPQSSRSFSVQTGDLTITNVQSDVGYYICQLINVAGSIITKAYLE	444
QY	1370	ILDPPQVPQTLEDIRALLAATGNPLPSVLTSPIGTOLVLDPGNSALLGCPKIGHPVPNIT	1429
Db	445	VTD-----VIADRP--PVIROGPNQTVAVD--GTFVLSCVAIGSPVPTIL	487
QY	1430	WFHGGPIVTAATGLTHHILAAGIOLQVANSLSGSGSQSEFSCLAQNEAGVLMQKASLVIQDY	1489
Db	488	WRKDGVLVSTQDSRIKQL---ENGVLQIRYAKLGDTRYCTIASTPSGEATWSAYIEVQFE	545
QY	1490	WWSVDRLATCASCNGRNGVQOQRLRCLLNSTEVNPAHCAGKVRPAQVPTA-----	1539
Db	546	-----GVVPQPR-----PTDNLIPSAPS-KPEVTDVSRNVTLSWOP	583
QY	1540	-CNRDCCPSRWMTSWSACT-----RSCGGGVTRRTVCOQLK-----ASGLS	1581

```
Query Match      2.9%; Score 277.5; DB 1; Length 802;  
Best Local Similarity 22.0%; Pred No.6.8e-13;  
Matches 138; Conservative 69; Mismatches 170; Indels 251; Gaps   39;
```

	QY	278	KIRNSGADSTVQIFFYOPIIHRWRETDFPCSATCGGGYGOLTSACCYDLRSNRVVAD--	335 : : :
Dd	370	KIRELTSLDHP-QSPFVDPE-----	--GSIKLVA-----RVVLERI 403 :	
QY	336	----QYCHYYYPENI-----KPKPKLOECLNDPCPADSGYKIIMPYDLVHPLPRWEATP	384 : : : : : :	
Dd	404	ARKEQCNFVPDNIDDIADVLAPEEKEDDTPTETCIYSN-----W-SP	445 : : : : : :	
QY	385	WTACSSS-CGGGTQSRAVSCEEDIQHVIIVSEWKMCYTTPKMPIAQPCTFDCKPWLAQ	443 : : :::	
Dd	446	WSACSSTSCEGRMMR-----QRMLKAQLDLISVP-----CPD-TQ	479 :: : :	
QY	444	EWSPTVTCCOGLRYRYVYLICDHIRMHTGCCPSKTKPHIEECIVTPTCYKPKEKLVEA	503 :: :	
Dd	480	DFQPC---MGPG-----CSDEQG-----STCMMS-----	500 : :	

```
Best Local Similarity 22.0%; Pred. No. 6.8e-13;
Matches 138; Conservative 69; Mismatches 170; Indels 251; Gaps 39;

QY 278 KIRNSGSADSTVQIFIPYQPIIHRWRETDFFPCSATCGGQVLTSAECYDLRSNRVAD-- 335
Db 370 KIRPLTSLDHP-QSPFYDPE-----GGSIKIVA-----RVVLERI 403
QY 336 ----QYCHYPENI-----KPKPKLQECNLDPCPASDGYKQIMFYDLYHPLPRWEATP 384
Db 404 ARKGEQCNFVNDIDIVADLAFEEKEEDDTPTCYISN-----W--SP 445
QY 385 WTACSS--GGGTOSRAVSCVEEDIOGHVTSVEEWMKMTTPKMPIAOPCNFIDCPKMLAQ 443
Db 446 WSACSSSTCEKGRMR-----ORMLKAQLDLSVP-----CPD--TQ 479
QY 444 EWSPTCTCGVGLQTRDVFCSHLLSRENNETVILADELCRQPKPSTVQACNRFNCP-----P 730
Db 480 DFQPC---MGPG-----CSDEGD-----STCMMS----- 500
QY 504 KLPWFKAQAELEBGAAYSEFPFPEAWSACTVTCGVGTQVRIVRQCVLLSFSQSVDLP 563
Db 501 --DWI-----TWSPCSVSCGMGTRSR-----ERYVKQFP 527
QY 564 ID--ECGPKPASORACYAGPCSGE---IPFNPDETDGLFGQLQDFELYDWEYEGFTK 618
Db 528 EDGSMCKVPTETEKIVNEECSPSCIVTEWG-----EWD-----E 564
QY 619 CSCEGCGGVQEAUVVSCNLKQTRPAENLC--VTSRRPQLLKSCNLDPC--PARWEIGKW 675
Db 565 CSASCGTGKRR--HRMIKMT--PADGSMCKAETTEAKCMMPCHTIPCLLSPW--SEW 618
QY 676 SPCSLTCGVGLQTRDVFCSHLLSRENNETVILADELCRQPKPSTVQACNRFNCP-----P 730
Db 619 SDCSVTCGKGRTR-----ORMLKSAEELGD--CNE-ELEQAEKCMLEPCPIDCELT 667
QY 731 AWYPAQWQPCSRFCGGGVQKREVLCQRMADGSLFLPETF-----C----- 772
Db 668 EW--SOWSECNISCGKGMIRTRMIKIEPQFGG-TACPETVQTKCRVRKCLRGPGMEKR 724
QY 773 ----SASKPAQOACKDD-----CP-SEWLLSDWTECSTSCGEGTQTSACRKLMT 821
Db 725 RWKEAREKRRSEQAANKNIDNEQYPVCRLEPW--TAWTECSTLCGGGIQERYMMVKRRSKS 782
QY 822 GLSTVNVSTLCPLPFSRSSIRPCMLATC 849
Db 783 -----TQFTSCKD---KKELRACNVHPC 802

RESULT 14
US-08-313-288B-12
; Sequence 12, Application US/08313288B
; Patent No. 5750502
; GENERAL INFORMATION:
; APPLICANT: Jessell, Thomas M. and Avihu Klar
; TITLE OF INVENTION: CLONING, EXPRESSION AND USES OF A
; TITLE OF INVENTION: NOVEL SECRETED PROTEIN, F-SPONDIN
; NUMBER OF SEQUENCES: 20
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Cooper & Dunham LLP
; STREET: 1185 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10036
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/313,288B
; FILING DATE: January 5, 1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: White, John P.
; REGISTRATION NUMBER: 28,678
; REFERENCE/DOCKET NUMBER: 40028-A-PCT-US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 278-0400
; TELEFAX: (212) 391-0526
; TELEX:
; INFORMATION FOR SEQ ID NO: 12:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 802 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-313-288B-12

Query Match 2.9%; Score 277.5; DB 1; Length 802;

PCT-US93-03164-12
; Sequence 12, Application PC/TUS9303164
; GENERAL INFORMATION:
; APPLICANT: Jessell, Thomas M
; APPLICANT: Klar, Avihu
; TITLE OF INVENTION: CLONING, EXPRESSION AND USES OF A
; TITLE OF INVENTION: NOVEL SECRETED PROTEIN, F-SPONDIN
; NUMBER OF SEQUENCES: 20
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Cooper & Dunham
; STREET: 30 Rockefeller Plaza
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10112
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US93/03164
; FILING DATE: 19930402
; CLASSIFICATION:
```


GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 24, 2002, 04:17:47 ; Search time 45.82 Seconds
(without alignments)
3695.098 Million cell updates/sec

Title: US-10-044-807-2

Perfect score: 9588

Sequence: 1 MECCRRATPTGTLILFLAFL.....LKLQLSQPKRCRCGTCGKA 1762

Scoring table:

BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283138 seqs, 96089334 residues

Total number of hits satisfying chosen parameters: 283138

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

PIR-71.*

1: pir1.*

2: pir2.*

3: pir3.*

4: pir4.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1145.5	11.9	2165	2 T21371	hypothetical prote
2	1088.5	11.4	1059	2 T22545	hypothetical prote
3	887	9.3	1558	2 C89114	protein C37C3.6a f
4	887	9.3	2167	2 T34395	hypothetical prote
5	732.5	7.6	951	2 T00260	hypothetical prote
6	630.5	6.7	1205	2 T18517	procollagen N-endo
7	597	6.2	951	2 T00017	gene ADAMTS-1 prot
8	584.5	6.1	550	2 T47158	hypothetical prote
9	566.5	5.9	898	2 T14764	hypothetical prote
10	468.5	4.9	1444	2 T18856	angiogenesis inhib
11	422	4.4	837	2 T00355	hypothetical prote
12	373	3.9	788	2 T25061	hypothetical prote
13	372	3.9	4391	2 A38096	perlecan precursor
14	352.5	3.7	5175	2 T20992	hypothetical prote
15	352.5	3.7	5198	2 T43290	hemichitin precurs
16	341	3.6	957	2 T15976	hypothetical prote
17	332.5	3.5	1906	1 S68235	myosin-light-chain
18	326	3.4	3707	2 S18252	heparan sulfate pr
19	320.5	3.3	1259	2 S36126	neural cell adhesi
20	318.5	3.3	7962	2 T38436	elastic titin - hu
21	317.5	3.3	6642	2 T29757	protein UNC-89 - hu
22	317	3.3	1584	2 T00026	brain-specific ang
23	312.5	3.3	1572	2 T00027	brain-specific ang
24	311.5	3.2	469	1 S29126	properdin precursor
25	311.5	3.2	1447	2 A54100	tumor suppressor p
26	309.5	3.2	4162	2 T42633	connectin/titin -
27	307.5	3.2	1260	1 S05479	neural cell adhesi
28	302.5	3.2	1239	1 A32579	neuroglian - fruit
29	301	3.1	1427	2 I51669	tumor suppressor -

ALIGNMENTS

RESULT 1

T21371

hypothetical protein F25H8.3 - Caenorhabditis elegans

C:Species: Caenorhabditis elegans

C>Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 29-Oct-1999

C:Accession: T21371; T24896

R:Gajadaty, S.

submitted to the EMBL Data Library, February 1996

A:Reference number: Z19413

A:Accession: T21371

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-2165 <WIL>

A:Cross-references: EMBL:Z69360; PIDN:CAA93287.1; GSPDB:GN00022; CESP:F25H8.3

A:Experimental source: clone F25H8

R:Gajadaty, S.

submitted to the EMBL Data Library, February 1996

A:Reference number: Z19949

A:Accession: T24896

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-2165 <W12>

A:Cross-references: EMBL:Z69361; PIDN:CAA93288.1; GSPDB:GN00022; CESP:F25H8.3

A:Experimental source: clone T13H10

C:Genetics:

A:Gene: CESP:F25H8.3

A:Map position: 4

A:Introns: 31/1; 52/1; 135/2; 193/3; 216/1; 266/1; 495/2; 547/3; 584/3; 634/2; 744/1;

Query Match

Best Local Similarity 11.9%; Score 1145.5; DB 2; Length 2165;

Matches 414; Conservative 185; Mismatches 553; Indels 723; Gaps 74;

QY	33	DGLDWDANGPWSCESTCGGASYSRLRLS-----SKSCEGRNIRYTCNSVDCPPPEAG	86
DB	603	DQOWDWSGSCSTCGGVQKGLRDCSPKPRNGKVCVGQRYRSCNTQECQPDWQ	562
QY	87	DFAOQCSAHN--DYKHKG--QFWEWLP--VSNDPDNPSCLKQAKGTTLVVEAPKVL	139
DB	663	PYREVQCSEFNKNDIGIOGVASTNTHTWPKYANAPNERCKLYCRLSGSAFYLLRDKVV	722
QY	140	DGTRCYTESLDMCISGLCQIVGCDHQLGSTVKEDNCVCNGDGSTCRLVRGQYKQLSAT	199
DB	723	DGTFCDRNGDDICVAGACMPAGCDHQLHSTLRDKCGVCGDDSSCKVVKGTENEQ--GT	780
QY	200	KSDDTVAIPVGSRHRLVLKG-----PDHLYLETKTQGTGKGENSLSSGTGFLVDSNV	254
DB	781	FGYNEVMKIPAGSANIDIRQKGYNNMKEDDNYL---SLRAANGFELLNGHQVSLAQOI	837
QY	255	DFQ-----KFPDKILRM--AGPLTADFIYKIRNSSADSTVOFIF-----YQPI	297

neural cell adhesi
neural cell adhesi
sax-3 protein - Ca
Bravo/Nr-CAM cell
hypothetical prote
properdin - mouse
neural cell adhesi
sdk protein - frui
neural cell adhesi
semaphorin F precu
rig-1 protein - mo
transmembrane rece
zonadhesin - mouse
connectin 3B - chi
dutt1 protein - mo
transient axonal g

Db 838 AFQDTVLEYSGDALIERINGTPIRSDIYVHVLSVSGSHPIDISYEYMTAAVNAVIRPI 897
QY 298 ---IHRRETD-FPPCSATCGGGQTLISAECYDILRSNRVYADQCHYYPENI-KPKPKLQ 352
Db 898 SSALYLRVTDWTECDRAC-RGQOSKMLCLMSTHROSHDRNC---QNVLPKQATR 952
QY 353 BCNLD-----PCPASIGYKQ-----IMPYDLYHPLP----- 378
Db 953 MCNIDCSTRWITEDVSSCSAKCGSQKQKRVSCVKMEGDRQTPASEHLCDNRNKFSDIAS 1012
QY 379 -----RWEATPWTACSSCG-GGQSRASVSCVEEDIQGHVTSVEEMKCMYTPKMPIA 429
Db 1013 CYIDCSGRKNYGEWTSCTCGSGKMKRKSVCVDDSNR-----RVDESICGREQKEATE 1068
QY 430 OPCNIFDCPKWLAQWSPCTVTCGGGLRYRVVLCID--IRGMHTGGCSP-KTKPHIKEBC 486
Db 1069 RECNRIPCRWYIGHWSECSRCDGVMRHAQCLDAADREHTTTRCGPAQTOEHCNEHA 1128
QY 487 IVPTPCYKPKELPVEAKLPWFQKQAELEEGAAVEEPSFIPEAWSACTVTCGVGTQVRI 546
Db 1129 CT-----WWQFG-----VMSDCSAKCGDGQVYRD 1152
QY 547 VRCOVLLSFQSQVADLPIDCEGPKPASQACRAGYPCSGEIEPEFNDETDLGLFGQLDFD 606
Db 1153 ANCT-----DRHRSVLPHEHRLKMEKLIITKPCHRESC-----PKY----- 1187
QY 607 ELYDWEYEGFTKCSGCGGVQAEVYVCLNKQTPREAEENLCVTSRRPOLLKSCNLDPC 666
Db 1188 KLGEW-----SQGSVCEDGWSRRYSVCVSGNGTE--VMSLCTGASDRPASHQTCNLGTC 1241
QY 667 PARWEIGKSPCLTCGVLQTRDVFCSHLLSREMNETVILADELCRQPK-PSTVQACNR 725
Db 1242 PF-WRNTDWSACSVCGIGHRETEC---IYREQS---VDASFCGDTKMPETSQCHL 1293
QY 726 FNGPPAPYPAQWOPCSTCGGVQKREVLCKQRMADGSLFELPFTCSA-SKPACQOACK 784
Db 1294 LPC-TSMKPSHWSPCSVTCGGIOTRSVSC--TRGSEITVD--EYFCDRNIRPLKKTCE 1349
QY 785 KDC-----PSEWLLSDWTEGSTCGEGTQTRSAICRKLMTGLSTVYVNST 830
Db 1350 KDCDGPVLQKLDQADVPPIRATGPTACSGTNGTQRLKCRDHVD-----LPDE 1404
QY 831 LCPLPSSSIRICMLATPCRPSTKSPHIAAARKVYIOTRRORLKHFFVVGGFAYLL 890
Db 1405 YCNHLDKEVTRNRLDCS-----Y 1425
QY 891 PTAVALRCPARVRKPLITWEKDGQHLISSTHTV--APGYLKIHLKPSDAGVYCS 948
Db 1426 WKMAEWEECPA-----TCGTHVQOSRNVTCVSAEDGGRTI--LKDVDCDVQ--- 1469
QY 949 AGPAREHVIKIGNKRKLVARPLS-----PRSEEVLAGRKGPKKEALQTHKHQ 999
Db 1470 -----KRPTSARNCRLEPCPKGEHI-----GS 1492
QY 1000 GIFNGSKAEKRLGAANPSRYDOLVSRLLBQGGWPGELLASWEAQDSAEANTTSEEDPG 1059
Db 1493 WLGWNSK-----SASCG-----GGW-----RRRSVCTSSS 1520
QY 1060 AEQVLLHLPFTWTEQRKLDLILGNLSOPEELRDLISKULYAQAELFRSHLEHQDTL 1119
Db 1521 CDE-----TRKPKMFD----- 1531
QY 1120 LKPSERTSPVT-----LSPKHVHSFSSSLTSTGDAGGSGRRPRKPTILRKISAAQ 1174
Db 1532 -KCNEELCPPLTNNSWQISPTWTHCS-----VSCGGGVQV-----RKIWCED 1571
QY 1175 QLSASEVTHLQTVALASGTLVLLHCEAIGHPRPTTSWANGEVQFSDRILLQPDPS 1234
Db 1572 VLS-----GRQDDIEGSE---IKPRE- 1590
QY 1235 LQILAEVADGVFYTCNATNALGYDSVATVLAGKPLVKTSMVTINTEKPAVTVDIGS 1294
Db 1591 -----QRDCMPPCR----- 1600

QY 1295 TIKTVGVNVTINCOVAGVPAEVTWERNKSKLGSPHLLHSGSILLTNVSSSDOGLYSR 1354
Db 1601 -----SHYKNTSSASAKTSLSSNSNTSSA 1626
QY 1355 AANLHGELETSTOLLILDPQPQVPTOLEDIRALLATGPNLPVSLTSPILGTQLVLDPGNSA 1414
Db 1627 S-----ASSLPILP----- 1635
QY 1415 LLGCPKIGHPPVNTWTFHGGQPIVATATGLTHLLAAGQILQVANLSSGGSGEESCLAQNE 1474
Db 1636 -----PVVS----- 1639
QY 1475 AGVLMOKASLVLDYVWSVDRLATCSACSGNRGVQOQPRCLLNSTEVNPA--HCAGKVR 1532
Db 1640 -----WQTSAWSACSACG-RGTRKRVVVCYNPSLNVTVASTECQDTTK 1682
QY 1533 PAVOPIACNRDPCSRWMTVTSWACTRSGGGVQTRRVTCQKLKASGISTPVSNDMCTQV 1592
Db 1683 P-VVEVRCRTKHP-RWKTTWSSCSVTGGRIRREVOCYR---GRKNLVSDESQNPX 1736
QY 1593 AKR-----PVDQACNOOLCVENAFSSWGQCGPCIGPHILAVQHVQFCOTRDGITLP 1645
Db 1737 TKLNSVANCFFV---ACP---AYRNVTWPSKCKDEARGQ--KQTRVHCISTSGKRAA 1788
QY 1646 SEQCSALPRPVSTQNCWSEACSVHWRVSIWTLCTATCGNYGFSQRRVREC----- 1694
Db 1789 PRMCELARAPTSIRECDTSCNCPYEWVPGDWQTCSSKSGE-GVQIREVRCRRKINFNSTIP 1847
QY 1695 --VHARTNKAVPHRLCSWGPANWQRCNITPCENMBCRDTRYCEKVKQLKCOLSOQFK 1752
Db 1848 IIFMLEDEPAVPKREKCELPKPNESQTCELNPCD-----SEFK 1895
QY 1753 -----SRCCGTCK 1761
Db 1886 WSGPWGECSSKNCQ 1900

RESULT 2
T22545
hypothetical protein F53B6.2 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
C:Accession: T22545
R:White, S.
submitted to the EMBL Data Library, October 1996
A:Reference number: Z19578
A:Accession: T22545
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-1059 <WIL>
A:Cross-references: EMBL:Z81086; PIDN:CAB03121.1; GSPDB:GN00019; CESP:F53B6.2
A:Experimental source: clone F53B6
C:Genetics:
A:Gene: CESP:F53B6.2
A:Map position: 1
A:Introns: 38/3; 92/3; 131/3; 169/1; 236/2; 282/3; 349/2; 376/2; 420/1; 576/3; 759/3;

Query Match 11.4%; Score 1088.5; DB 2; Length 1059;
Best Local Similarity 19.3%; Pred. No. 1.4e-54;
Matches 350; Conservative 188; Mismatches 446; Indels 827; Gaps 50;

QY 11 TLLFLAFLLLSRTPARSEDRDGL-WDANGPKSECSRTCGGASYSRLRCLSSKSCGR 69
Db 6 TVVFLFYLLIASFHV-----DALSWAASPWSCTKTCGGGVSRQLRCLTSK-CSGE 58
QY 70 NTRYRCPNSWDCPEPEADFRQAQCSAHNDVKHHGQYFVWLPVSNPDNPCLSKCAQKGT 129
Db 59 SVRFKCAOKTCKSKRLARDTTCGG-EIIVSRGQ-----CEVVCRLTG 103
QY 130 LVVELAPKVLDTGRCY-TESLDMCISGLCOIVGCDHQLGSTVKEDNCNGCNGDSTCLV 188

Db 104 --ANLWRVDDGTPCOAATSAVCSKSGQIVGCDGLISSFRFDACVCGGRGDTIC--- 158
 QY 189 RGOYKSQLSATKSDDDTVVAIPYGSRIHRLVLKGPDLHYLETKTLOGTKGENSLSTGIFL 248
 Db 159 ----- 158
 QY 249 VNSSVDFOKFPDKELLRMAGPLTADFIKIRNSGSAOSTVQFIFQPIIHRRETFDEPP 308
 Db 159 -DNKG-----FIWKV----- 308
 QY 309 CSATC-----GGGYLTSAE-----CYDLRSNRVADQYCHYTPENIKPKPKQECNL 356
 Db 174 CASNDDDIVDWSGAGASTASTSOPIVVCVNAITGRVPEKLC---ADKLRPKVEARPCPM 230
 QY 357 DPASADGYKOIMPYDLYHPLPRWEATPWTACSSCGGIGIOASRVCE--EDIOGHVTS 414
 Db 231 LICPS-----RWAAOWTECVPHCGEGTRKREYICVQTAHNVTVH--- 270
 QY 415 VEEWKCTYTPKMP IAOPCNIFDCPKWLAQWSPCVVTTCGGLRYRVVLCIDHRGMHTGG- 473
 Db 271 VPDTCFENGTRPAABENCVSTSCGWEAGKSKCTASCQGVRRHVACV----- 323
 QY 474 -CSPKTRPIKEECIVPPPCYKPKLEPVEAKLPWFOAELEGAAYS-----EPPS 525
 Db 324 CDDEGRPRQETTCYAGIPCSIATN-----SLDNDRAY-LOGNTFGSDMNDHNDWQAPR 376
 QY 526 PIPEAWSACTVTCGVGTQVIRVRCQVLLSFSOSVADLPIDCEG-PKPASORACYAGPCS 584
 Db 377 LVAGEWSTCSSTCGTGYMSRIVECVAVNPISAPIKIPMSECQDQEQKLFESCEVRS- 435
 QY 585 GEIPEFNPDETDLFGGLQDPDELYDMYEGFTKCSFSCGGGVOEAVVCSLKNKOTREPAP 644
 Db 436 -----PLQED-----SKLSEDEAPYQWRYGDTQCSASCLGGKKAALKCIQVSTGKSVQ 485
 QY 645 ENLCVTSRRPOLLKSNLDPARPWIGKWSQSLFCGVGLQTRDYFCSHLLSRE--MN 702
 Db 486 WSOCDARRRPPKSRPCNQHPQPPFWITSKYSDCMSGSGTARRSVKCAQTVSKTDGAD 545
 QY 703 ETVILADELGRQPKPTQVQACNRFNCPPAWY-----PAWQPCSRCTGG 746
 Db 546 AHIVLRDRCHERKPKQETETCNVACPATVWSLKNRKNKIKLNKLTAKTQWTECSRSDS 605
 QY 747 GVKREVLCORMADGSFELPETFCSA-SKPACQOACKKDDCPSEWLLSDWTECSTSCG 805
 Db 606 GERROVWCEIRDSRGKTORRPDVECDANKP----- 637
 QY 806 EGTQTRSAICRMKLTGLSTVYNSTLCPLPFSSSRIPCMATCARPGRPSTKHSPHIA 865
 Db 638 -----QIVFVCSFGSCSRPE-----LL 654
 QY 866 ARKVIQTRRQKLHFVVGGEFAYLLPCTAVLRCPARRRKPLITWEKGOHLISSTHVT 925
 Db 655 SNRVFEQNAEQKLTILGIGVATLYGTSIKCPAKKFTDKKIYKWKNGKIKNDIAHK 714
 QY 926 VAPFGVLTIRLKPSPDAGVYTCAGAPREHFVTKLGGNRKLVARPLSPRSEEVLAGRK 985
 Db 715 VSANGLRVFARMEDAGYEC----- 736
 QY 986 GGPKEALQTHKHQNGIFSNKGAERGLAANPGSRYDDLVSRLLQGGWPGELLASWEAQ 1045
 Db 737 ----- 736
 QY 1046 DSAERNITSEEDPGAEQVLLHLPFTMTVTEQRRLDDILGNLSQOPELRDLYSKHLVAQLA 1105
 Db 737 -----FT-----DRLOQNVIT----- 746
 QY 1106 QEIFRSHLEHODTLKPSERKTSPTVLSPHKHVSGFSSSLRSTSTGAGGSRPRHRKPT 1165
 Db 747 -----LNFYRDF-----PASR----- 758
 QY 1166 ILRKISAAQLSASVTVHLQGTVALASGTLVLLHCEAIGHPRPTISWARGNEVEQFSD 1225
 Db 759 -----VDLA-----PKPQIPSTNRQVRQVSK 780

QY 1226 RILLOPDDSLQILAPVADVGFYTCNATNALGYDSVSIATVLAGKPLVKTRMTVINTX 1285
 Db 781 EDVLRQASV----- 790
 QY 1286 PAVTVDIGSTIKTVQGVNVNTINCOVAGVPEAEVTFWFRNKSGLGSPHHLHRSGLLTNVSS 1345
 Db 791 -----LHK-----MNVSL 798
 QY 1346 SDOGLYSCRAANLHGETESTQLLIDPPQVPTQLEDIRALLAATGPNLPSVLTSPGQTQ 1405
 Db 799 -----IEALL--TAPN----- 807
 QY 1406 LVLDPCNSALLGCPKIGHPVNPITWPHGGOPIVTATGLTHHTLAAGQIILQVANLSCSG 1465
 Db 808 ----- 807
 QY 1466 EFSCLAQNEAGVLMOKASLVIOQYMWSDVRLATCSASCGNRGVQOQPLRCLLNSTEVNPA 1525
 Db 808 -----DEKAREOLRKY-----GNELV----- 823
 QY 1526 HCAGKVRPAVQPIACNRDRPCSRMVTWSAC--TRSCG--GGVQTRVTCQKILKASGISTP 1583
 Db 824 -----ARWDIGHWSECRQKTVHAGVQARGISC-KVTFHGEIRN 861
 QY 1584 VSNDMCTOVAK-RPVDTQACNQOLCVENAFSSWGOCNGP-CIGPHLAYQVHROVFCQTRDG 1641
 Db 862 VDNSTICESIASVRPPETRPCHREDCPRWEASQWSECSQRCVSSMLAOKRRNVTCRTFNG 921
 QY 1642 ITLPSEQSALPRPVSTONCNSACSVHVRYSLWTLCTATCGNYGFSQRRRVECVHARTNK 1701
 Db 922 TSDVQIHCIDIINRATIMDCPNQCKAEWRTSDWGSCEGCTGGVGLRLLSVWISSGR 981
 QY 1702 AVPEHLCSWGRPPANWQRC-----NITPC-----ENMECRDTTRYCEKVKQLKQL 1748
 Db 982 PAGRN-CEQMRPHSARACVADEPLPCMPPTASALYQRDASCQOQSFCDIILKLFHSCDS 1040
 QY 1749 SOKSRCCGTC 1759
 Db 1041 LEVRQKCCSTC 1051

RESULT 3

C89114
 protein C37C3.6a [imported] - Caenorhabditis elegans
 C:Species: Caenorhabditis elegans
 C:Date: 10-May-2001 #sequence_revision 10-May-2001 #text_change 10-May-2001
 C:Accession: C89114
 R:anonymous, The C. elegans Sequencing Consortium.
 Science 282, 2012-2018, 1998
 A:Title: Genome sequence of the nematode C. elegans: a platform for investigating bio
 A:Reference number: A75000; MUID:99069613; PMID:9851916
 A:Note: see websites genome.wustl.edu/gsc/C.elegans/ and www.sanger.ac.uk/Projects/C-
 A:Note: published errata appeared in Science 283, 35, 1999; Science 283, 2103, 1999;
 A:Accession: C89114
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-1558 <STO>
 A:Cross-references: GB:chr_V; PIDN:AAC25867.1; PID:g3294501; GSPDB:GN00023; CESP:C37C
 C:Genetics:
 A:Gene: C37C3.6a
 A:Map position: 5

Query Match

Best Local Similarity 9.3%; Score 887; DB 2; Length 1558;
 Matches 229; Conservative 107; Mismatches 274; Indels 178; Gaps 29;

QY 27 RSEEDRDGLWAWGPMWSECSRTCGGASYSLRCLSSKSKSECRNIRYRTCSNVDCPPENG 86
 Db 70 KSGQKETGNWGFVPENECSRSGGVQLEKRCQ--SGDCTGASVRYISCNLNAC--ESG 125
 QY 87 -DFAQQCSAHNDVKHGHQFYEWLFPVSNPDNPNCSLKCOAKGTTLVLVELAPKVLDRICY 145

C;Accession: T00260
R;Nagase, T.; Ishikawa, K.; Miyajima, N.; Tanaka, A.; Kotani, H.; Nomura, N.; Ohara, O.
DNA Res. 5, 31-39, 1998
A;Title: Prediction of the coding sequences of unidentified human genes. IX. The complete
A;Reference number: Z14086; MUID:98290545
A;Accession: T00260
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1-951 <NAG>
A;Cross-references: EMBL:AB011177; NID:g3043733; PIDN:BAA25531.1; PID:g3043734
A;Experimental source: brain
C;Genetics:
A;Note: KIAA0605
C;Superfamily: thrombospondin type 1 repeat homology
F;46-106/Domain: thrombospondin type 1 repeat homology <THRI>

[illegible]

RESULT 6
T18517
procollagen N-endopeptidase (EC 3.4.24.14) I - bovine
N:Alternate names: procollagen N-proteinase
C:Species: Bos primigenius taurus (cattle)
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
C:Accession: T18517
R:Collage, A.; Nusgens, B.V.; Lapiere, C.M.
submitted to the EMBL Data Library, February 1996
A:Description: Cloning of the cDNA of the bovine procollagen I N-proteinase.
A:Reference number: z18941
A:Accession: t18517
A:Status: preliminary; translated from GB/EMBL/DBBJ
A:Molecule type: mRNA
A:Residues: 1-1205 <COL>
A:Cross-references: EMBL:X66389; NID:e990769; PID:e228215; PID:CAA65253.1
C:Experimental source: skin
C:Genetics:
A:Gene: PC I-NP
C:Function:
A:Description: catalyzes cleavage of the propeptides of type I and II collagen
C:Keywords: hydrolase; metalloproteinase

Query Match	6.7%	Score 639.5;	DB 2;	Length 1205;
Best Local Similarity	22.8%	Pred No. 1.1e-28;		
Matches 208; Conservative 96; Mismatches 264; Indels 345; Gaps 34;				
QY	32	RDLWDANGPNWSECRTGGGASYSLRC-----LSSKSGEGRNIRYTCSNVDCDPEEA	85	
Db	554	RDGNWGANSFGSCSRTCCTGVKFETROCDNPHPANGORTCGSLAYDFQLCNSQDCPDAL	613	
QY	86	GDFRAQQCS-----AHDNVKHGHGFYEWLPVSN-DPDNPCSLCAOKAGTTLVVVELAPKV	138	
Db	614	ADFEEOCRQMDLYFHEHDAQBH-----WLFPHERDAKERCHLYCESKETGEVVYSMKRMV	668	
QY	139	LDGTRC-YTESILDMCISGLCOIGCDHQLGSTVEDNCVCNGDGOSTCRIVRGOYK---S	194	
Db	669	HGTRCSYKDFAFLCVRGDCRKVGCGDVIGSSGDKCGVGGGDNSCHKVVKGTFSRPX	728	
QY	195	QLSATKSDDTVVAIPYGSRRHLRLKGPDLHYLETKTLOGTK-----GENSL--STGTFLV	249	
Db	729	KLGVIK----MFEIPAGARALLIOEADTTSHLLAVKNLETGKFIINENDVDNPSKTFFIA	784	
QY	250	DNSSVDFOKFPDKELLRMAGPLTAFTVKLRNSGSADSTVOQIFPYQDIHRWRETFOPPC	309	
Db	785	MGVEWEYFDEGRETQTGMPLHGTTVLVPIPEGDARIS---LTYYMIHE-----832		
QY	310	SATCCGGYQILTSAECYDLRSNRVVADQYCHYPENIKPKPLECNLDPCPASDYGKQIM	369	
Db	833	-----DSUNVDNDNNVLED-----SVGRE---851		
QY	370	PYDLYHPLPRWEATPTACSSCGGGIQSPRAVSCVBEDTQGHTVSYEWNKMOMYTPKMPIA	429	
Db	852	-----WALKKUSPCKPCGGSGQTFKYGC-----875		

QY 430 QPCNIFDCPKWLAQEWSPCTVTCGGRLRYRVVLICIDHRGHTGSCPKTKPH-IKEECIV 488
DB 876 -----RRR-----LDHKMVRHGFCDSVSKPAIRRTC-- 902
QY 489 PTPCYKPKELPVEAKLPWFQKQALERGAASVEEPSFIPAWSACTVTCG-VGTQVRIV 547
DB 903 -----NPOE-----CSQPVVVTGEWEPSCRSRSGRTGMQVRVS 934
QY 548 RCQVLLSFSSQSVADLPIDCEGPKPASORACYAGPCSGEIPFENFDEIDGLFGGLQDFDE 607
DB 935 RC----- 936
QY 608 LYDWEYEGTKESCGGVQAEVAVVCLNKQTRPAEENLCVTSRRPPOLLKSCNLDPCP 667
DB 937 -----VOPLHNNTTRSVHYTKHCNDAR--PGRBRACNRELCP 970
QY 668 ARWEIGKWSPCSLTCGVGLQTRDVFCSHLLSREMNETVILADE---LCRQPKSTVOACN 724
DB 971 GWRAGSWSQCSVTGCGNGTOERPVLCR-----TADDSFGVCHEERPETARICR 1018
QY 725 RPNCP-----PAWYPAQW-----OPCSRT-----CGGGVOKREVLCCKORMADGSFL 765
DB 1019 LGPCPNTSDPSKSKSVVQWLSRDPNSPVQETSSKGRQCG---DKSVFCR-----M 1067
QY 766 ELPEPFCASAKPACQACQACKDCPCSEWLLSDWTECSTGCGEGTQTRSACRKLMTGLST 825
DB 1068 EVLSRYCSI--PGYNKLCCKSCNPHNL-----TDVDRAPPPSGKHNDEELMPTLSVPT 1121
QY 826 VNSTLTCPP-----LPFSSIRCMATCARPKRSTK-----HSPHIAA 865
DB 1122 LYMEVQPPGIPGLEVPLNTS-----STNATEDHPETNAVDVPYKIPGLEDEVPNNLIP 1175
QY 866 AR-KVYIOTRRQR 877
DB 1176 RRPSPYEKTRNQ 1188
RESULT 7
gene ADAMTS-1 protein - mouse
C:Species: Mus musculus (house mouse)
C:Date: 22-Jan-1999 #sequence_revision 22-Jan-1999 #text_change 21-Jul-2000
C:Accession: T00017
R:Kuno, K.; Lizasa, H.; Ohno, S.; Matsushima, K.
Genomics 46, 466-471, 1997
A:Title: the exon/intron organization and chromosomal mapping of the mouse ADAMTS-1 gene
A:Reference number: Z14055; MUID:98110583
A:Accession: T00017
A:Status: preliminary; translated from GB/EMBL/DDBJ
A:Molecule type: DNA
A:Residues: 1-951 <KUN>
A:Cross-references: EMBL:AB001735; NID:g2809056; PIDN:BAA24501.1; PID:g2809057
A:Experimental source: strain 129SVJ
C:Genetics:
A:Gene: ADAMTS-1
A:Introns: 228/1; 343/3; 388/1; 444/1; 539/3; 602/1; 660/3; 719/2
C:Superfamily: thrombospondin type 1 repeat homology
F:542-598/Domain: thrombospondin type 1 repeat homology <THR3>
Query Match 6.2%; Score 597; DB 2; Length 951;
Best Local Similarity 29.1%; Pred. No. 2.2e-26;
Matches 143; Conservative 51; Mismatches 180; Indels 118; Gaps 16;
QY 34 GLWDAMGPMWSECSRTCCGGGASYSRLRC-----LSSKSCGGRNIRYR 86
DB 544 GSKGPMGPMGDCSRTCCGGGQVYTMRECDNPNPKNGKYGCEGRVYRSCNIEDCPDNNK 603
QY 87 DPAQCCSAHNDVKH----HGQFYEWLP--VSNNDPNCSLKCAQKTTLVVELAPKVLID 140
DB 604 TFEEOCEAHNEFSKASFGNEPTVEVTPKYAGVSPKDRCKLCEAKGICGYFFVLQPKVD 663
QY 141 GTRCYTESLDMCTISGLCQIVGCDHQLGSTVYKEDNCVNGDGTCTCLRVGRQYKSQLSATK 200

DB 664 GTPCSPDSTVQVQGVQKAGCDRIIDSKKFKDKGCGGNGSTCKKMSG---IVTSTRP 720
QY 201 SDDIYVAIPYGRHTR-----LVLGKPDHLYLETWLTQTCGNSLSSTG 245
DB 721 GYHDIVTIPAGATNIEVKHRNQSRNNGSFLAIRAADGTIYINGNFTLSTLEQDLTYKG 780
QY 246 TEL-VDNSSVDQKFPDKKEILRMAGPLTADFIVKIRNSGSA-DSTVOIFYQYPIIHRWRE 303
DB 781 TVLIRSGSSAALER-----IRSFSLKPLTLQIOMVGHALPKPKIFYFM----- 826
QY 304 TDFPFCSATCGGYQLTSAECYDLRSNRVADQYCHYYPENIKPKPKLQECNLDPCPASD 363
DB 827 -----KKKTESFNAIP----- 837
QY 364 GYKQIMPYDLYHPLPWEATPWTACSSCGGTOGRVAVSCVEEDIOGHVTSVEWKMCT 423
DB 838 -----TFSEWVTEEWGECSTKCGSGWQRRVQC--RDINGHPAS-----ECALE 879
QY 424 PKMPIAOPCNIFDCPKWLAQEWSPCTVTCGGRLRYRVVLICIDHRG--MHTGGCSPKTKP- 480
DB 880 VKPASTRPCADLPCHPWQGDWSPCKTCKGKYKKRTLKCVSHDGGVLSNESCDPLKKPK 939
QY 481 HIKEECIVPTPC 492
DB 940 HYIDFCTL-TQC 950
RESULT 8
T47158
hypothetical protein DKFZp762C1110.1 - human (fragment)
C:Species: Homo sapiens (man)
C:Date: 20-Apr-2000 #sequence_revision 20-Apr-2000 #text_change 20-Apr-2000
C:Accession: T47158
R:Blum, H.; Bauersachs, S.; Mewes, H.W.; Weil, B.; Wiemann, S.
Submitted to the Protein Sequence Database, March 2000
A:Reference number: Z24379
A:Accession: T47158
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-550 <AAA>
A:Cross-references: EMBL:AL162080
A:Experimental source: adult melanoma (Mewo cell line); clone DKFZp762C1110
C:Genetics:
A:Note: DKFZp762C1110.1
Query Match 6.1%; Score 584.5; DB 2; Length 550;
Best Local Similarity 29.4%; Pred. No. 5.6e-26;
Matches 148; Conservative 59; Mismatches 168; Indels 129; Gaps 19;
QY 23 SRTARSEEDR--DGLWDAMGPMWSECSRTCCGGGASYSRLRC-----LSSKSCGGRNIRYR 74
DB 130 NKTDRKHEDTPPHGSGWGMGPGDCSRTCCGGGQVYTMRECDNPNPKNGKYGCEGRVTR 189
QY 75 TCSNVDCPEAG--DFRAOCCSAHNDVKH----HGQFYEWLP--VSNNDPNCSLKCAQK 127
DB 190 SCNLEDCPDNNKGTFFEOCEAHNEFSKASFGSGPAVEWIPKYAGVSPKDRCKLCAQK 249
QY 128 TTLVVELAPKVLDTGRCYTESLDMCTISGLCQIVGCDHQLGSTVYKEDNCVNGDGTCTCL 187
DB 250 IGYFFVLQPKVVDGTPCSPDSTSVCGVQGVQCVKAGCDRIIDSKKFKDKGCGGNGSTCKK 309
QY 188 VRGQYKSQLSATKSDDTVVAIPYGSRHRLVGLKPDHLYETWLTQTCGNSLSSTG 243
DB 310 ISG---SVTSAPGYHDIITPTGATNIE-----VKQRNQRGSRNNGSFLAKAA 356
QY 244 TGTFLVDNSSVDQKFPDKKEILRMAGPLTADFIVKIRNSGSA-DSTVOIFYQYPIIHRWR 302
DB 357 DGTYY-----LNGDYTL-----STLEQDIIMYKGVVLR-- 383
QY 303 ETDFPFCSATCGGYQLTSAECYDLRSNRVAD-----OYCHYPENI 345

```
Db 384 -----YSGSALERIRSFSPKLEPLTIOVLTVGNALRPKIKYIFV----- 425
QY 346 KPKPKLOECNLDPCPASDGYKIMPYDLYHPLPRWEATPWTACSSCGGGIQSRVSCVE 405
Db 426 --KKKESFNALP-----TFSAWVIEENGECSCGELGWQRLLVEC-- 464
QY 406 EDIQGHVTSVEBWKMYTPKMPIAQPCNIFDCPKWLAQEWSPCTVTCGGGLRVKAVLCID 465
Db 465 RLINQOPAS-----EKAKEVKFASRPRCADHPCPQWOLGWSKCTGKGGKRSKLCLIS 520
QY 466 HRG--MHTGGCSPKTKP-HIKEEC 486
Db 521 HDGGVLSHESCDPLKKPKHFIDFC 544

RESULT 9
T14764
Hypothetical protein DKF2p434H204.1 - human (fragment)
C:Species: Homo sapiens (man)
C>Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 20-Sep-1999
C:Accession: T14764
R:Wambutt, R.; Heubner, D.; Mewes, H.W.; Gassenhuber, J.; Wiemann, S.
submitted to the Protein Sequence Database, August 1999
A:Reference number: Z18181
A:Accession: T14764
A>Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-898 <WAM>
A:Cross-references: EMBL:AL110226
A:Experimental source: adult testis; clone DKF2p434H204
C:Genetics:
A:Note: DKF2p434H204.1

Query Match 5.9%; Score 566.5; DB 2; Length 898;
Best Local Similarity 23.3%; Pred. No. 1.1e-24;
Matches 279; Conservative 118; Mismatches 404; Indels 397; Gaps 63;

QY 608 LYDWEYEGFTKSCSGGGVQAVVSCLNKQTRPAENLVCVTSRRPQLLKSNLDPCP 667
Db 34 VFSWHYGPWTKCTVTCGRGVQNVYCLERQA-GPVDEEHCDPLGRPDQQRKCEQPCP 92
QY 668 ARWEICKWSPCLTCGV-GLOTRDFVCSHLLSRENETVILADELCRO-PKPSTVQACNR 725
Db 93 ARWAGIEWQJCSGSGGGLSRRAVLG--IRSVGLDEQSALEPPACEHLPRPTETPCNR 150
QY 726 -FNCPPAWYPAQWQPCSRCTGGGVQKREVLCKORMADGSFLELPETFC-SASKPACQAC 783
Db 151 HVPCTPATWAGVNNNSQSVTCGEGTQRNVLCTNDTC-----VP---CDEAQPAASEVTC 201
QY 784 KKDCPSEWLLSDWTECSTSCGGTQTSRAICRKMVLKTLGLSTVWNSLTCP-PLPSSSIR 842
Db 202 SLPLC--RWPLG--TLGPEGSGSSSHELF-----NEADFIPHLAPRSP-ASSPK 249
QY 843 PCMIATCARCPRTSKHSPHIAAARKVYIOTRQRKLFHVVGFAYLPLKTAVALRCPAR 902
Db 250 PGTW-----GNAIEEAEPLDLPFGVF-----VDDFYF----- 277
QY 903 RVRKPLITWEKGQHLISSHTVIVAFVGYLKTIRLKPSSDAGVYTCGAPAREHVIKLI 962
Db 278 -----DYNFINPHE-----DLSYGPSEEPDLDLACT 303
QY 963 GNRKLV--ARPLSPRSEEVLAGKGPKEALQTHKHQNGIFSNGSKA---EKRLAAMP 1017
Db 304 GDRTPPHSRPAAPSTGSPVATEPPAAKE-----EGVLGPWSPSPWSPQAGRSPPP 355
QY 1018 GSRVYDLVSLRLGGWGPGLLASWEAQDSARNNTTSEEDPCAEQVLLHLPF----- 1069
Db 356 PS-----EQ--TPGNFLI-----NFTPEETPGADPLGLPSLWPRVST 393
QY 1070 ----TWYEQRRLLDILG--NLSCQPEELRDLYSKHLVAQLAQEIFRSHLEHQDTLLKPS 1123
Db 394 DGLQTPATPESQNDFFVQKDSQSLPPPPWRD-----RTNEVFKDDDEPK----- 437
```

```
QY 1124 ERTSPVTLSPKHVSGFSSSLRTSSGCDAGGSRPRPKPTILRKLSAAQQLSASEVVT 1183
Db 438 -----GRGAPHLPP-----RFSSTLPPPLSP----- 457
QY 1184 HLGOTVALASGTLVYLHCEAIGHPRPTISW--ARNGEEVQFSDRIILQPDQSLQILAVP 1241
Db 458 -VGSTHSSPSPDVAFL-----WTGCTVAVPALEGG-----LGPV 491
QY 1242 EADVGFYTCNATNALGVDYSIAVTLAGKPLVKTSMVTINTEKPAVTVDI-----GS-T 1295
Db 492 DSEL-----WPTGVGA-SLLPPPIAPLPEMKVDRDSSLEPCTSPFPAPGSGWD 538
QY 1296 IKTVGVNVTINCOVAGVPEAEVTFWRNKSGLSPHHLHSGSLLLNVSSDQGLYSCLA 1355
Db 539 LQTV-----AVM-----GTFLPTLT-----GL----- 556
QY 1356 ANHGLTESTOLLILDPQVPTOLEDIRALLAANTGNPLSVLTSLGTLQVLDPGNSAL 1415
Db 557 ----GHMPEPA--LMPGKGPQESLS-----PEV--PLSSRLSLTPAWDS- 593
QY 1416 LGCPKIGHVPVNTWFGGGPIVITATGLTHHILAAGQILOVANLSSGSGQGEFSCLAQNEA 1475
Db 594 ----PANSRVRPET-----QPLAPSLA-----EA 613
QY 1476 GVLMOKASLIVDYWWSVDRLATCSACGNGVQOPRLRCLLNSTEVNPAHCAGKVRPAV 1535
Db 614 G--PPADPLVVRNASQWAGNSECSTTCGLGAVWRP-VRCSSGRDE---DCAPAGEP-- 664
QY 1536 QPT-ACNRDCPSRMWTSWACTRSCGGGVQTRRVTCQKLKASGIST-----PVSNDMCT 1590
Db 665 OPARRCHLRPC-ATWHSNWNKSCRSOGSSVRDVQC-----VDTRDLRLRPFHCQ 716
QY 1591 QVAKRVEDTQACNOQLCVWEAFSSWGQCGNCPICGPHLAVQHVQECQTRDGTILPSEQCS 1650
Db 717 PGPAKPPAHPCCAQPCLSWITSSHRESEACGG---GEQRLVTC-PEPGL-----CE 766
QY 1651 ALPRPVSTONCSEACSVHVRVSLWTLTATCGNVGFSRVECVCHARTKNAVPEH--L 1707
Db 767 EALRPNTTTPCTHPTC-QWVVGPMWQCSAPCGG-GVORRLVKVNTQT--GLPEEDSDQ 822
QY 1708 CSMGPPANWQCNITPCENME---C---RDTTRYCEKVKQKLCQKQFSKRCQCGTC 1759
Db 823 CGHEAMPSSRRCGTDCPEVPEPRCERDRLSFGCEITLRLGRCOLPTINTQCRSC 880

RESULT 10
T18856
angio genesis inhibitor homolog - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C>Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 18-Feb-2000
C:Accession: T18856; T24653
R:McMurray, A.
submitted to the EMBL Data Library, July 1995
A:Reference number: Z19031
A:Accession: T18856
A>Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-1444 <WIL>
A:Cross-references: EMBL:Z50004; PIDN:CAA90293.1; GSPDB:GN00028; CESP:C02B4.1
A:Experimental source: clone C02B4
R:McMurray, A.
submitted to the EMBL Data Library, July 1995
A:Reference number: Z19917
A:Accession: T24653
A>Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-1444 <WIZ>
A:Cross-references: EMBL:Z50006; PIDN:CAA90302.1; GSPDB:GN00028; CESP:C02B4.1
A:Experimental source: clone T07C5
C:Genetics:
A:Gene: CESP:C02B4.1
A:Map position: X
```

RESULTS

A:Accession: A41059
A:Molecule type: mRNA
A:Residues: 'RT',892-908,'R',910-1101,'L',1103-1132,'L',1134-1221,'L',1223-1397 <KAS>
A:Cross-references: GB:S76436; NID:g243370; PIDN:AAB21121.1; PID:g243371
R:Dodge, G.R.; Kovalsky, I.; Chu, M.L.; Hassell, J.R.; McBride, O.W.; Yi, H.F.; Ioza-Genomics 10, 673-680, 1991
A>Title: Heparan sulfate proteoglycan of human colon: partial molecular cloning, cell
A:Reference number: A40306; MUID:91365376
A:Accession: A40306
A:Molecule type: mRNA
A:Residues: 1018-1405,'G',1407-1409,'G',1411-1465 <DOD>
A:Cross-references: GB:M64283; NID:g184424; PIDN:AAA52699.1; PID:g184425
J:Heremans, A.; van der Schueren, B.; De Cock, B.; Paulsson, M.; Cassiman, J.J.; van
J. Cell Biol. 109, 3199-3211, 1989
A>Title: Matrix-associated heparan sulfate proteoglycan: core protein-specific monoclonal
anes.
A:Reference number: A33625; MUID:90078352
A:Accession: B33625
A:Molecule type: protein
A:Residues: 1379-1384,'X',1386-1388,'X',1390-1398 <HE2>
A:Accession: A33625
A:Molecule type: protein
A:Residues: 2166-2171,'X',2173-2175,'X',2177-2185 <HE3>
A>Note: peptide potentially matches four different regions of sequence shown
C:Genetics:
A:Gene: GDB:HSPG2
A:Cross-references: GDB:I26372; OMIM:142461
A:Map position: lp36.1-lp36.1
C:Superfamily: LDL receptor ligand-binding repeat homology; EGF homology; laminin G r
F:1-21/Domain: chondroitin sulfate proteoglycan; glycoprotein; heparan sulfate; transmem
F:22-4391/Product: perlecan #status predicted <SIG>
F:22-193/Domain: I <DOM1>
F:194-530/Domain: II <DOM2>
F:199-234/Domain: LDL receptor ligand-binding repeat homology <LDL1>
F:285-319/Domain: LDL receptor ligand-binding repeat homology <LDL2>
F:325-359/Domain: LDL receptor ligand-binding repeat homology <LDL3>
F:368-403/Domain: LDL receptor ligand-binding repeat homology <LDL4>
F:531-1676/Domain: III <DOM3>
F:1159-1206/Domain: laminin-type EGF-like homology <LEG>
F:1563-1610/Domain: laminin-type EGF-like homology <EG7>
F:1613-1668/Domain: laminin-type EGF-like homology <LEG8>
F:1677-3686/Domain: IV <DOM4>
F:2007-2034/Domain: transmembrane #status predicted <TRM>
F:3687-4391/Domain: V <DOM5>
F:3845-3880/Domain: EGF homology <EGF1>
F:3888-3921/Domain: EGF homology <EGF>
F:3953-4106/Domain: laminin G repeat homology <LG2>
F:4147-4175/Domain: EGF homology <EGF2>
F:4149-4151/Region: motor neuron attachment (L-R-E) motif
F:4299-4301/Region: motor neuron attachment (L-R-E) motif
F:65-7176/Binding site: heparan sulfate (Ser) (covalent) #status predicted
F:89,554,1755,2121,3072,3105,3279,3780,3836,4068/Binding site: carbohydrate (Asn) (co
F:2995,3933,4179/Binding site: chondroitin sulfate (Ser) (covalent) #status predicted

Query Match 3.9%; Score 372; DB 2; Length 4391;
Best Local Similarity 20.9%; Pred. No. 1.4e-12;
Matches 358; Conservative 178; Mismatches 559; Indels 618; Gaps 96;

QY 75 TCSNVDCPPEAGDFRAQCASHNDVKHHQGQFFEWLPVSNDDPNPCS-----LKCAKGW 128
||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 838 TCD--ACAPGYTGRCCSCAPG-----YEGNPFI--QPQCKRPVNQEIIVRCDEGS 884
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 129 -----TLVELAPRVLDGT-RCYTESLDMCTISGLC-----QIV 160
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 885 MGTSGEACRCNNVVGRLCNCADSGSFHLSTANDPDCLKCFCMGVSRHCTSWSRAQLH 944
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 161 GCDHQLG-----STVKEDNCGVN-----GDGSTCRVLRGQYKSLSATKSDTVVA 207
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 945 GASEEPGHFSLTNAASTHTTNEGIFSTPPGELGSSFHRLSLGGPYFWSLPSRFGLGDKVTS 1004
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 208 IPYGSRRIR-----LVUKDPDHYLETPTKLQCTKGENSLSTSTGTFLL 248

Db 3206 QTTDVEVL-----VPPRIEDERVLOGKEGNTYVHVCQVTPVPTVTKR--- 3252
QY 1001 IFSNGSKAEK-----RGLAANPGSRYDDLYSLRLLGOGWPGPELLASWEAQDSARNIT 1053
Db 3253 ---NGKEIQPNPVLHNRATRADEGKYCIASN--EAGTAVADFIDVFTKPTFETHE 3307
QY 1054 SEEDPGAQVLLHL-----PFT-----MVTEQRRLDDILG 1083
Db 3308 TFNIVEGESAKIECKIDGHPKPTISWLKGGPFNMNIIILSPRGDTLMILKAQRFD--G 3364
QY 1084 NL-----SQPEELRDLYSKHLVAQLAQAEIFRSHLEHODTLTKPSERTSPVTL 1132
Db 3365 GLYTCVATNSYGDSEQDFKV--NVYTKPIDETIDQTPKAVAGEIILKCPVLGNPTPTVT 3423
QY 1133 -----SPKHVGSFSSLSRTSSTGDAG-----GGSRRPH-----RKPTI 1166
Db 3424 WKRGDDAVPNDSRHTVNNYDLKINSVTTEDAGQYSCIADVNEAGNLTHYAAEVIGKPTF 3483
QY 1167 LRKISAAQQLSASEVTHLQGTVALASGTLVLLHCEAIGHPRPTTISWARGVEVQSDR 1226
Db 3484 VRKGGNLYEVIENDTIT-----MDCGVTSRPLPSISWFRGDKPVLVYDR 3527
QY 1227 ILLOPDDS--LOILAPVEADVGYTCNATNALGYDSVSIATVTLAKGLVKTSRMTVINTEK 1285
Db 3528 YSISPDGSHITINKAKLSDGKYICRASNEAGTSDIDLILKLVPPKIDKSN----- 3580
QY 1286 PAVTVDIGSTIKTVQGVNVTINCOVAGVPAEVTWFRNKSGLSPHHLHEGSLILT--- 1341
Db 3581 -----IGNPLAIV--ARTIYLECPISGIPQPDVITWTKN---GMDINMTDSRVILAQNE 3629
QY 1342 -----NVSSSDQGLYSCRAANLHGETESTQILLIDPPO-----V 1376
Db 3630 TFGIENVVOTDQGRVCTATNRRGRASHUDFLDVLSPPEFDHNGTOPTIKREGDTITLTC 3689
QY 1377 PQQL-----EDIAL-----LAATGPNL-----PSVLTSPLGTOLVLD 1409
Db 3690 PIKLAEDIADVMVSWTKDSRALDGLTDNVDISDGRKLTISQASLENAGLYTCIALN 3749
QY 1410 PGNSA-----LLGCPKIGHPVNITWFGHQPTV 1438
Db 3750 RAGEASLEFKVEILSPVIDISRNDVQVAVNQVPTIMRCVATGTHFFPSIKLNGKEVT 3809
QY 1439 TATGLTHILAAQTLQVANT--SGSQGEFFCSIAQNEAGVLMQKASLVITODYWMSVDRL-- 1496
Db 3810 DDENI--RIVEQGVQVILRTSDSHAGKWSVCAENDAGV--KELEMVLDFVTPPVVSVKS 3865
QY 1497 -----ATCSAGNGVQVQVPLRLCLLNSTEVPNPAHCAKVRPAVQPIACNRERDCP-- 1546
Db 3866 DNPICALGETITLFCNAGSNYPOLKWKAGGSLIFDPSDGARI-----SLKAGRLDIPHL 3920
QY 1547 SRMVTWSACTRSCGGGVQTRRTYCKLAKAGISTPYSNDMCTQVAKRPVDTQACNOOL 1606
Db 3921 KKTVDGY-----TCQALNAG-----TSEASVSVDV----- 3947
QY 1607 CVEWAFSSWGCNCPGCPGPHLAVOHQVFCOTRDI-----TLPSE-----OCSALPRPVS 1657
Db 3948 -----LVPEET-----NRDGDMSPLPQAQSLTLOCLAGQKPPV 3982
QY 1658 TQNCWSEACSVHVRVSLWTLCTATCG 1683
Db 3983 -----QMRWTLNGTALTHTSPG 3999

RESULT 15
T43290
hemiscentin precursor - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 11-Jan-2000 #sequence_revision 11-Jan-2000 #text_change 18-Feb-2000
C:Accession: T43290; T20993; T24734
R:Vogel, B.E.; Hedgecock, E.M.
submitted to the EMBL Data Library, June 1998
A:Description: Hemiscentin is required for hemidesmosome mediated cell adhesion and germ-
A:Reference number: 222396

A:Accession: T43290
A:Status: preliminary; translated from GB/EMBL/DBRJ
A:Molecule type: mRNA
A:Residues: 1-5198 <VOG>
A:Cross-references: EMBL:AF074901; PIDN:AAC26792.1
R:Sulston, J.
submitted to the EMBL Data Library, December 1994
A:Reference number: Z19355
A:Accession: T20993
A:Status: preliminary; translated from GB/EMBL/DBRJ
A:Molecule type: DNA
A:Residues: 1-5198 <WIL>
A:Cross-references: EMBL:Z47068; PIDN:CAA87336.1; GSPDB:GN00028; CESP:F15G9.4b
A:Experimental source: clone F15G9
R:Kershaw, J.
submitted to the EMBL Data Library, December 1994
A:Reference number: Z19929
A:Accession: T24734
A:Status: preliminary; translated from GB/EMBL/DBRJ
A:Molecule type: DNA
A:Residues: 1-5198 <W12>
A:Cross-references: EMBL:Z47070; PIDN:CAA87345.1; GSPDB:GN00028; CESP:F15G9.4b
A:Experimental source: clone T09B9
C:Genetics:
A:Gene: him-4; F15G9.4b
A:Map position: X
A:Introns: 85/1; 120/1; 334/3; 370/1; 477/2; 606/3; 664/1; 935/3; 977/1; 1051/3; 1184/1; 2512/2; 2593/3; 2699/3; 2759/1; 2852/1; 2889/3; 2913/3; 2941/1; 2967/3; 2991/3; 3031; 4225/1; 4361/1; 4408/1; 4456/1; 4498/1; 4647/3; 4838/1; 4902/1; 4964/1; 5034/1; 51

Query Match 3.7%; Score 352.5; DB 2; Length 5198;
Best Local Similarity 19.1%; Pred. No. 2.3e-11;
Matches 188; Conservative 138; Mismatches 329; Indels 331; Gaps 38;

QY 897 LRCPARRRVKKITWEKOGHLISSTHTVAPPGYKLIHRLKPSDAGVTC-----SAGPA 952
Db 3146 IRCEVFNPEPVNWLKDGEPYISDQLQSTKLSYLHRETTLADGGTTCATNKAGES 3205
QY 953 REHFVKLGGNKKLVARPLSPR--SEEEVLAKRGK-----GPKALQTHKHONG 1000
Db 3206 OTTQDEVL-----VPPRIEDERVLOGKEGNTYVHVCQVTPVPTVTKR--- 3252
QY 1001 IFSNGSKAEK-----RGLAANPGSRYDDLYSLRLLGOGWPGCELLASWEAQDSARNIT 1053
Db 3253 ---NGKEIQPNPVLHNRATRADEGKYCIASN--EAGTAVADFIDVFTKPTFETHE 3307
QY 1054 SEEDPGAQVLLHL-----PFT-----MVTEQRRLDDILG 1083
Db 3308 TFNIVEGESAKIECKIDGHPKPTISWLKGGPFNMNIIILSPRGDTLMILKAQRFD--G 3364
QY 1084 NL-----SQPEELRDLYSKHLVAQLAQAEIFRSHLEHODTLTKPSERTSPVTL 1132
Db 3365 GLYTCVATNSYGDSEQDFKV--NVYTKPIDETIDQTPKAVAGEIILKCPVLGNPTPTVT 3423
QY 1133 -----SPKHVGSFSSLSRTSSTGDAG-----GGSRRPH-----RKPTI 1166
Db 3424 WKRGDDAVPNDSRHTVNNYDLKINSVTTEDAGQYSCIADVNEAGNLTHYAAEVIGKPTF 3483
QY 1167 LRKISAAQQLSASEVTHLQGTVALASGTLVLLHCEAIGHPRPTTISWARGVEVQSDR 1226
Db 3484 VRKGGNLYEVIENDTIT-----MDCGVTSRPLPSISWFRGDKPVLVYDR 3527
QY 1227 ILLOPDDS--LOILAPVEADVGYTCNATNALGYDSVSIATVTLAKGLVKTSRMTVINTEK 1285
Db 3528 YSISPDGSHITINKAKLSDGKYICRASNEAGTSDIDLILKLVPPKIDKSN----- 3580
QY 1286 PAVTVDIGSTIKTVQGVNVTINCOVAGVPAEVTWFRNKSGLSPHHLHEGSLILT--- 1341
Db 3581 -----IGNPLAIV--ARTIYLECPISGIPQPDVITWTKN---GMDINMTDSRVILAQNE 3629
QY 1342 -----NVSSSDQGLYSCRAANLHGETESTQILLIDPPO-----V 1376
Db 1376

Query Match	1.1%;	Score 56.4;	DB 1;	Length 2277;
Best Local Similarity	44.9%;	Pred No. 0.00041;		
Matches 256;	Conservative	0;	Mismatches 311;	Indels 3; Gaps 1;
2936	aggtcttcggggaggaagggcgccggaagagggccctgcagaccacaacaccaga	2995		
530	AGGAGATCAAGGGCGCACCGTGGCCGTGGACTGGGCGGTGGCCAGNCAGTACAGG	589		
2996	acggggatcttctcaacggcgagcaagggcgagaaagcggggcctggccgccaccacccgggga	3055		

ATTORNEY/AGENT INFORMATION:
NAME: Osman Ph.D., Richard A
REGISTRATION NUMBER: 36,627
REFERENCE/DOCKET NUMBER: UCB96-055
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415)343-4341
TELEFAX: (415)343-4342
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 2277 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: CDNA
US-09-098-487-5

Query Match 1.1%; Score 56.4; DB 2; Length 2277;
Best Local Similarity 44.9%; Pred. No. 0.00041;
Matches 256; Conservative 0; Mismatches 311; Indels 3

	Matches	256;	Conservative	0;	Mismatches	311;	Indels	3;	Gaps	1;
QY	2936	aggtgcttgqgggaggaaagcgccgcgaaggagggccctcagaccacaaaaccaga	2995							
Dd	530	AGGAGATCAAGGGCCGCACCGTGGCTGGACTGGGCCGTGGCCAAAGACAAGTACAAGG	589							
QY	2996	acgggatcttctcaacaggcagcaagcggaagcggggcctgccgcacaaccgggga	3055							
Dd	590	ACACCCAGAGCGTGAGCGCCATCGCGCAGGAGAAGACCACAGCAAGCACCCAGGAGA	649							
QY	3056	gcgcgtacgagaccttgtctcccggctgtgtgagcaggcggtggtccccggagagtgc	3115							
Dd	650	GCGTGAAGAAGAGGGCCGCAGGAGGAGGACATGGAGGAGGAGGACACGACGACG	709							
QY	3116	tggccctctgtgggggcgcaggacttcgggaaaggaacacgaacctcgagaggaacccgg	3175							
Dd	710	ACACACAGCAGCAGGAGGAGCGGCCGTGTTCGACGACGAGGACGAGGAGGAGAACA	769							
QY	3176	gtgcagagcaagtgtctctgtcacctgcccttcacctgtgtgacgcagcagcgcgcctgg	3235							
Dd	770	TCSAGAGCAAGGTGACCAAGCCGTCGAGATCCAGAAGCGCGCGTGAAGCGCCCCGCC	829							
QY	3236	aqaacatctctggggaaacctctcccagcagcccgaggagctgtgcgcaccttacagaagc	3295							
Dd	830	CGGCCAAGAGCAGCAGCACACAGCGAGGAGGACACGCACCTGGAGGAGAGCGCACAGCATCG	889							
QY	3296	acctggtgcccagctggcccaggagattcttcggagccaaccttyggaccaggacacgc	3355							
Dd	890	ACGACGCCGAGAGCTTTGGCCC---AGAGGGACACACAGCACCCGAGGAGCAGGAGACAAGG	946							
QY	3356	tctgaagccctcggagcgcaggacttccccagtactctctgcctcataaacagtgt	3415							
Dd	947	CCGTCCAGGTGAGCAACANGAAGACCGCAAGCTGCCCCAGCGACGTGAACGAGGGCAAGA	1006							
QY	3416	ctggcttcagcagctctcttcgggaacctctccaccgggacgcggggaggtcttcgaa	3475							
Dd	1007	CCGTGTTTCATCGCAACCTGAGCTTCGACAGCAGGAGGAGGAGGAGCTGGCGGAGCTGCTGC	1066							
QY	3476	ggccacacgcgaagcccaccatccctgcga	3505							
Dd	1067	AGCAGTTCCGGCGAGCTGAAGTACGTGCCGA	1096							

RESULT 4

```

US-08-125-468-1
; Sequence 1, Application US/08125468
; Patent No. 5589385
; GENERAL INFORMATION:
; APPLICANT: Ryan, Michael J.
; APPLICANT: Lotvin, Jason A.
; APPLICANT: Strathy, Nancy
; APPLICANT: Pantini, Susan E.
; TITLE OF INVENTION: Cloning of the biosynthetic pathway for
; chlorotetracycline and tetracycline formation and cosmids

```

```

1  TITLE OF INVENTION:  useful therein
2
3  NUMBER OF SEQUENCES:  1
4  CORRESPONDENCE ADDRESS:
5  ADDRESSEE:  American Cyanamid Company
6  STREET:  One Cyanamid Plaza
7  CITY:  Wayne
8  STATE:  New Jersey
9  COUNTRY:  USA
10
11  ZIP:  07470
12
13  COMPUTER READABLE FORM:
14  MEDIUM TYPE:  Floppy disk
15  COMPUTER:  IBM PC compatible
16  OPERATING SYSTEM:  PC-DOS/MS-DOS
17  SOFTWARE:  PatentIn Release #1.0, Version #1.25
18
19  CURRENT APPLICATION DATA:
20  APPLICATION NUMBER:  US/08/125,468
21  FILING DATE:  22-SEP-1993
22  CLASSIFICATION:  435
23
24  ATTORNEY/AGENT INFORMATION:
25  NAME:  Tsevdos, Estelle J
26  REGISTRATION NUMBER:  31,145
27  REFERENCE/DOCKET NUMBER:  31,255-02
28
29  TELECOMMUNICATION INFORMATION:
30  TELEPHONE:  (201)831-3241
31  TELEFAX:  (201)831-3305
32
33  INFORMATION FOR SEQ ID NO:  1:
34  SEQUENCE CHARACTERISTICS:
35  LENGTH:  30001 base pairs
36  TYPE:  nucleic acid
37  STRANDEDNESS:  single
38  TOPOLOGY:  linear
39
40  MOLECULE TYPE:  DNA (genomic)
41
42  US-08-125-468-1

```

Query Match 1.0%; Score 54.6; DB 1; Length 30001;
Best Local Similarity 45.6%; Pred. No. 0.0029;
Matches 235; Conservative 0; Mismatches 274; Indels 0

QY	2960	gccgaagaggccctgcagaccacaacacgaagcgggtcttctccaagcgagca	3019
Db	15355	gCGGATGTCGCCCTCAACACGGCTACTTCCAGGCGAAGCGGTGAGAGCGCGGTG	15414
QY	3020	aggcggagaagcgggctgcccgaacccgggagcgcgtacagacacctcgtctccc	3079
Db	15415	AGCTCGGCTCTTCAGCTGCTTCGCCGAGCGCTCCGCCGGGTCCACCAAGATCCGCGCG	15474
QY	3080	ggctgctggagcggcgcgctggcccggagagctgctggctcgtgggagcgcaaggact	3139
Db	15475	AAC TGCGGCGTCCGGACCGGCTGTTTCAAGGACTTCTGAAGCCCTGCTGGCCPCGCCC	15534
QY	3140	ccgcggaagaacacgacctcgaggaggaccgggtgcgagagcaagtgcctcgacc	3199
Db	15535	TGCTGGACGACGAGCGCGCTACCGGGGCTCCGAGTCCGCCGGAGTTCCTGCTCC	15594
QY	3200	tgccttcacatggtgaccgagcagcgcgccctggacacacatcctggggaacctctcc-	3258
Db	15595	CCGCCCCACGTACTCTGGCGGCACCGCCGCCAGCAGCTCGGCTGCATTCACACGCT	15654
QY	3259	-----cagcagccaggagctgcgcacctctcacgaagcacctggtggcccagctgg	3313
Db	15655	GGCGCAGCTGACCGAGCGCTGGCGACGCAAGGCCAAGTCGGCGCTGCCCGCGGAGG	15714
QY	3314	ccagggagatcttcgcagccacctggagaccagcacgctcctgaagccctcggagc	3373
Db	15715	GCCAGCTGGCCTACCCCAAGCAGTACGAGGACCTGAGCCGCGCCCGCAGATCATGTGC	15774
QY	3374	gcaggactccccagtgactctctgcctcataaacagctgtcgtgcttcagagctccc	3433
Db	15775	ACATGGACGCCCAACAGGTTTCAGGCCCAACGATTTGGCGCGCGGATCCACTGGAGCC	15834
QY	3434	tgcgagacctctccacccgggagcgccgggggagcg	3468


```
Db 15835 GGTACACCTCTTGTGGAGCTCGGGGGGGGGC 15869
RESULT 5
US-08-474-933-1
; Sequence 1, Application US/08474933
; Patent No. 5866410
; GENERAL INFORMATION:
; APPLICANT: Ryan, Michael J.
; APPLICANT: Lotvin, Jason A.
; APPLICANT: Strathly, Nancy
; APPLICANT: Pantini, Susan E.
; TITLE OF INVENTION: Cloning of the biosynthetic pathway for
; TITLE OF INVENTION: chlorotetracycline and tetracycline Formation and cosmid
; TITLE OF INVENTION: useful therein
; NUMBER OF SEQUENCES: 1
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: American Cyanamid Company
; STREET: One Cyanamid Plaza
; CITY: Wayne
; STATE: New Jersey
; COUNTRY: USA
; ZIP: 07470
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/474,933
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/125,468
; FILING DATE: 22-SEP-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Tsevdos, Estelle J
; REGISTRATION NUMBER: 31,145
; REFERENCE/DOCKET NUMBER: 31,255-02
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (201)831-3241
; TELEFAX: (201)831-3305
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 30001 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
US-08-474-933-1

Query Match 1.0%; Score 54.6; DB 2; Length 30001;
Best Local Similarity 45.6%; Pred. No. 0.0029;
Matches 235; Conservative 0; Mismatches 274; Indels 6; Gaps 1;

QY 2960 gccggaaggagccctgcagaccacacacagagggatctctcacaacgcagca 3019
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 15355 GCGGGATCGTCGCGCCCTCAACACCGCTACTTCCAGCGGAAGCGCGTCGACAGCGCGGTGC 15414
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 3020 aggcgagaagcgggctgtgcccacacccggggagcgcgtacgacgacctgtctccc 3079
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 15415 AGCTCGGCTCTTCGAGCTGCTGCGCGAGCGCTCCGCGGGCTCGACAGATCGCGCG 15474
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 3080 ggtctgtgagcaggcggctgtgccccggagagctgtgtgctcgttggagcgcagact 3139
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 15475 AACTGGGCTCGCGACCGCGTGTTCAGGACTTCTTGAACGCGCTGTGTCGCGCTCGGCC 15534
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 3140 ccgcgaaagaaacacacacacacacacacacacacacacacacacacacacacac 3199
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 15535 TGCTGGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGAC 15594
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 3200 tgcccttcaccatggtgaccagagcagcggcgtggagacacatcctctgggaacctctcc- 3258
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
```



```
; NAME/KEY: misc_feature
; LOCATION: 14771..15154
; OTHER INFORMATION: /function= "potential open reading
; OTHER INFORMATION: frame"
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 15154..15426
; OTHER INFORMATION: /function= "potential open reading
; OTHER INFORMATION: frame"
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 15429..15664
; OTHER INFORMATION: /function= "potential open reading
; OTHER INFORMATION: frame"
; US-08-508-004-3
```

Query Match
Best Local Similarity 1.0%; Score 53; DB 1; Length 15664;
Matches 137; Conservative 0; Mismatches 140; Indels 0; Gaps 0;

QY	3075	ctcccgctgctgagcaggggcgctgcgccggagagtgtgacctggtggaggcgca	3134
Dd	12078	CCCTGGACACTGTCGTTCAGCGCCTGTGCCGCACCTGTTCGATGGTGGCGCGCA	12137
QY	3135	ggactcgcggaaaggaacacgaacctcgaggaggaccgggtgcagaagaatgcttct	3194
Dd	12138	CGACTTCGAACAACGGCAGGGCGGCACCGGGCAGGAGCACCTTGACGCCGTGATTCCA	12197
QY	3195	gcacctgccccttcaccattgttgacggagcagcgcgctggagacaatactctgggaaacct	3254
Dd	12198	CDCGCTGGTACTGAATGGTTCCGAAAGCACCGGCCGCTGTCGATGACC GCCGACAC	12257
QY	3255	ctccccagcacccgagagctgcgcacctctacagaagcaacctggtgccagctggc	3314
Dd	12258	GGTAGCGCTTACCGAGCGCCTGCTGGACGCCGCCGAGCGCCCATGMAAGCGCAGAGGC	12317
QY	3315	ccaggagactttccgcagccaacctggagcaaccaggac	3351
Dd	12318	CGCCGAGTTACCGCCCGCCACCGAGGACGACCAGCACAGGAC	12354

RESULT

```
LOCATION: 6403..7770
OTHER INFORMATION: /function= "potential open reading
OTHER INFORMATION: frame"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 7770..8006
OTHER INFORMATION: /function= "potential open reading
OTHER INFORMATION: frame"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 8033..8236
OTHER INFORMATION: /function= "potential open reading
OTHER INFORMATION: frame"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 8244..9443
OTHER INFORMATION: /function= "potential open reading
OTHER INFORMATION: frame"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 9450..10244
OTHER INFORMATION: /function= "potential open reading
OTHER INFORMATION: frame"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 10371..10586
OTHER INFORMATION: /function= "potential open reading
OTHER INFORMATION: frame"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 11115..11786
OTHER INFORMATION: /function= "potential open reading
OTHER INFORMATION: frame"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 11917..12741
OTHER INFORMATION: /function= "potential open reading
OTHER INFORMATION: frame"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 12748..14499
OTHER INFORMATION: /function= "potential open reading
OTHER INFORMATION: frame"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 14771..15154
OTHER INFORMATION: /function= "potential open reading
OTHER INFORMATION: frame"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 15154..15426
OTHER INFORMATION: /function= "potential open reading
OTHER INFORMATION: frame"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 15429..15664
OTHER INFORMATION: /function= "potential open reading
OTHER INFORMATION: frame"
IS-08-402-066-3
```

	Query Match	1.0%	Score 53;	DB 1;	Length 15664;
	Best Local Similarity	49.5%;	Pred. No. 0.0055;		
	Matches 137;	Conservative	0;	Mismatches 140;	Indels 0; Gaps 0;
QY	3075	ctcccggctgctgagcagggcgctgcccggagagctgctgctctgtggaggcgca	3134		
Db	12078	CCCGTGCACCTGCTTCGATCGAGCCCTGTGCCGCACCTGTTCCGATGTTGGGCGGCGA	12137		
QY	3135	ggactccqcggaaggaacacgactcgaggagagaccgggtgcagagcagaagtgcctct	3194		
Db	12138	CGAGTTTCACAAACGGCGAGSGCGGCACCGGGCAGGAGCACTTGGACCGCTGATTTTCCA	12197		

QY	3195	gcacctgcccttcaccatggtgaccagcagcagcgcgcttgagcagacatcctcggggaacct	3254
Db	12198	 CGCGCTGGTACTAGAAATGTTCCGCAAGCACCGGCGCGCTGTTCCATGACCGCGCGGAACAC	12257
QY	3255	ctccagcagcccgagagctgcgcgaacctctacagcaagcacttggtggcccgactggc	3314
Db	12258	 GGTAGCCTTTACCGAGGCCCTGCTGGAGCGCGCGACGCCATGAAGCGCAAGAGCG	12317
QY	3315	ccaggagatcttcgcgaagccacctggagaccaggac	3351
Db	12318	 CGCGGATTACCGGCCCGGCCACCAAGAGCAGCAGCAGGAC	12354

```

RESULT 10
US-08-402-068-3
; Sequence 3, Application US/08402068
; Patent No. 5633159
; GENERAL INFORMATION:
; APPLICANT: Pearson, Robert E.
; APPLICANT: Dickinson, Julie A.
; APPLICANT: Hamilton, Paul T.
; APPLICANT: Little, Michael C.
; APPLICANT: Beyer Jr., Wayne F.
; TITLE OF INVENTION: MYCOBACTERIOPHAGE SPECIFIC FOR THE
; TITLE OF INVENTION: MYCOBACTERIUM TUBERCULOSIS COMPLEX
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Richard J. Rodrick, Becton Dickinson and
; ADDRESSEE: Company
; STREET: 1 Becton Drive
; CITY: Franklin Lakes
; STATE: NJ
; COUNTRY: US
; ZIP: 07417
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/402.068
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Fugit, Donna R.
; REGISTRATION NUMBER: 32,135
; REFERENCE/DOCKET NUMBER: P-3283
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 15664 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 222..425
; OTHER INFORMATION: /function= "potential open reading
; OTHER INFORMATION: frame"
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 451..747
; OTHER INFORMATION: /function= "potential open reading
; OTHER INFORMATION: frame"
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 747..1109
; OTHER INFORMATION: /function= "potential open reading
; OTHER INFORMATION: frame"
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 1109..2014
; OTHER INFORMATION: /function= "potential open reading

```

Wed Jul 24 11:35:35 2002

NAME/KEY: misc.feature
LOCATION: 11115..11786
OTHER INFORMATION: /function= "potential open reading"
OTHER INFORMATION: frame"
FEATURE:
NAME/KEY: misc.feature
LOCATION: 11917..12741
OTHER INFORMATION: /function= "potential open reading"
OTHER INFORMATION: frame"
FEATURE:
NAME/KEY: misc.feature
LOCATION: 12748..14499
OTHER INFORMATION: /function= "potential open reading"
OTHER INFORMATION: frame"
FEATURE:
NAME/KEY: misc.feature
LOCATION: 14771..15154
OTHER INFORMATION: /function= "potential open reading"
OTHER INFORMATION: frame"
FEATURE:
NAME/KEY: misc.feature
LOCATION: 15154..15426
OTHER INFORMATION: /function= "potential open reading"
OTHER INFORMATION: frame"
FEATURE:
NAME/KEY: misc.feature
LOCATION: 15429..15664
OTHER INFORMATION: /function= "potential open reading"
OTHER INFORMATION: frame"
US-08-402-068-3

Query Match 1.0%; Score 53; DB 1; Length 15664;
Best Local Similarity 49.5%; Pred. No. 0.0055;
Matches 137; Conservative 0; Mismatches 140; Indels 0; Gaps 0;

QY 3075 ctccggctgctggagcagggcggtggccggagagctgctggcctctgggagggcga 3134
Db 12078 CCGTGGCACCCTGCTTCGACGCCCTGTGCGGACCTGTTCGCATGTGTGGGGCGCGA 12137
QY 3135 ggactccgggaaagaaacacgacctcggagggagggaccgggtgcagagcaagtgcct 3194
Db 12138 CGAGTTTCGACACGGCGAGGGCGGCACCGGGCAGGACACCTGGACGCCGTGATTTCGA 12197
QY 3195 gacactgacctaccatgagccagcagcggcctgagcagacatcctggggaacct 3254
Db 12198 CGCGTGTGCTGAATGTTTCGCAAGCACCGCGCCGCTGTCATGACCGCGCAACAC 12257
QY 3255 ctcccgagcggcggagctgcgcgacctctacagcaagcactgggcccagctggc 3314
Db 12258 GGTAGCGCTTACCGAGGCCCTGCTGGACGCCGCCGAGAGGCCATGAAGCGCAAGAGGC 12317
QY 3315 ccagagagatctccgagaccactggagaccaggac 3351
Db 12318 CGCCGAGTTTACCCGCCGCCACCGAGGACGACAGGAC 12354

RESULT 11
US-08-537-002A-4
Sequence 4, Application US/08537002A
Patent No. 5773282
GENERAL INFORMATION:
APPLICANT: TSUSAKI, Keiji
APPLICANT: KUBOTA, Michio
APPLICANT: SUGIMOTO, Toshiyuki
TITLE OF INVENTION: RECOMBINANT THERMOSTABLE ENZYME FOR
TITLE OF INVENTION: CONVERTING MALTOSE INTO TREHALOSE
NUMBER OF SEQUENCES: 17
CORRESPONDENCE ADDRESS:
ADDRESS: BROWDY AND NEIMARK
STREET: 419 Seventh Street, N.W., Suite 300
CITY: Washington
STATE: D.C.

COUNTRY: USA
ZIP: 20004
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent in Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/537,002A
FILING DATE: 29-SEP-1995
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 260984/1994
FILING DATE: 01-OCT-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP No. 5773282 yet received
FILING DATE: 08-SEP-1995
ATTORNEY/AGENT INFORMATION:
NAME: BROWDY, Roger L.
REGISTRATION NUMBER: 25,618
REFERENCE/DOCKET NUMBER: TSUSAKI-1
TELEPHONE: 202-628-5197
TELEFAX: 202-737-3528
TELEX: 248633
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 2889 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
US-08-537-002A-4

Query Match 1.0%; Score 50.4; DB 1; Length 2889;
Best Local Similarity 48.0%; Pred. No. 0.012;
Matches 215; Conservative 0; Mismatches 221; Indels 12; Gaps 2;

QY 2956 ggcggccggcagagggccctgcagaccccaaacacacagagggatctctccaaagcc 3015
Db 1486 GACCTCCCTTGGAGGCTACCAAGGCTGCTCCCTGGAGCTCTTCGACGAAACCC 1545
QY 3016 agcaagcgaggaagcgggcctggcccaacccgggagcgccgctacgacgacctctc 3075
Db 1546 TTCCTCCCGGTGAGGGGCGCTACCGCTTGACCTGGGCCCCCAGCGCTTGGCCCTCTTC 1605
QY 3076 tcccggtgctggagcagggcggtggccggagagctgctggcctctgggagggcag 3135
Db 1606 GCCCTGAAGCCCGTGGAGGCGGTGCTCCACCTCCCTCCCGGCTGGGCGGAGAGCCC 1665
QY 3136 gactcccggaagaaacacgacctcggagagagagaccgggtgcagagcaagtctctg 3195
Db 1666 GCCCTCCGAGGAGCGGACCTCCCGGCTCCACATGCTCCCGGGGCGGAGGCTCTCTCTG 1725
QY 3196 caactgccttcacatggtgacagagcagcgcgccctggacacatcctgggaacctc 3255
Db 1726 G-----TGGACACCTTGTCCACGAAGGGGGCGGAGGAGCTCTTAACGCCCTC 1776
QY 3256 tccagcagcggcggagagctgcgcgacctctacagaaagacct---ggtggccagctg 3312
Db 1777 GCCCAGACCTTGAAGGAGAGAGTGGCTGCGCCCTCAAGCGCGAGAGGTGGCCCTCTG 1836
QY 3313 gcccaaggagatctcccgagccacctggagcaccagagcagcctctgaagccctcgag 3372
Db 1837 GACGCCCTCCGCTTCCAGAGAGGACCCGCCCTTTACCTCACCTCTGCTCAGCTGGAGAAC 1896
QY 3373 gcagagacttcccgagctactctcgc 3400
Db 1897 CACAGACCTCCAGGTCTCCCTCCCTCC 1924

RESULT 12


```

US-08-863-010-4
; Sequence 4, Application US/08863010
; Patent No. 6087146
; GENERAL INFORMATION:
; APPLICANT: TSUSAKI, Keiji
; APPLICANT: KUBOTA, Michio
; APPLICANT: SUGIMOTO, Toshiyuki
; TITLE OF INVENTION: RECOMBINANT THERMOSTABLE ENZYME FOR
; TITLE OF INVENTION: CONVERTING MALTOSE INTO TREHALOSE
; NUMBER OF SEQUENCES: 17
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: BROWDY AND NEIMARK
; STREET: 419 Seventh Street, N.W., Suite 300
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20004
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/863,010
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/537,002
; FILING DATE: 29-SEP-1995
; APPLICATION NUMBER: JP 260984/1994
; FILING DATE: 01-OCT-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP No. 6087146 yet received
; FILING DATE: 08-SEP-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: BROWDY, Roger L.
; REGISTRATION NUMBER: 25,618
; REFERENCE/DOCKET NUMBER: TSUSAKI-1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-628-5197
; TELEFAX: 202-737-3528
; TELEX: 248633
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2889 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; PS-08-863-010-4

```

Query Match	1.0%;	Score 50.4;	DB 3;	Length 2899;
Best Local Similarity	48.0%;	Prod. No. 0.012;		
Matches 215;	Conservative 0;	Mismatches 221;	Indels 12;	Gaps 2;
QY 2956	ggcgcccgagagggccctgcagaccacacacagagggatcttctccaagggc	3015		
Db				
1486	GACCTCCCTTGGAGGCTACCAAGGCTCTGCTCCCTGGAGCTTCTCGACGCAACCC	1545		
QY 3016	agcaagggcggaagcgggccctggccgccacacccggggagccgtactacgacactcgtc	3075		
Db				
1546	TTCCCCCGGTGGAGGGCGCTACCGCTTGACCTGGGCCCCACGGCTTCGGCCCTCTTC	1605		
QY 3076	tcccggctgtgcagcagggcggtggcccgagagctgtgccttgtggagggcagcag	3135		
Db				
1606	GCCTGAAGCCGTGGAGGCGGTGCTCACCTCCCTCCCCGACTGGGGCGAGGAGCC	1665		
QY 3136	gabtccgggaagaaacagcactcgaggagaccgggtgcagagcaagtctcctg	3195		
Db				
1666	GCCCCGAGGAGGCCGACCTGTCCCCGGGTCCATGCCCCGGGGGCGGAGGTCTCTGT	1725		
QY 3196	cactggccttcacatggtgacagagcgcgctggacgacatcctgggaaacctc	3255		

```

Db 1726 G-----TGGACACCCCTGGTCCACGAAGGGGGGAGGAGCTCCTTAACGCCCTC 1776
QY 3256 tccacgagcccgagagctgcgcgacctctacagaagcacct---ggtggcccgctg 3312
Db 1777 GCCCAGACCCTGAAGGAGAAGAGCTGGCTCGCCCTCAAGCCGAGAAAGGTGGCCCTCCTG 1836
QY 3313 gccacgagagatctccgcagccacctggagcaccaggaagcagcgtctctgaagccctcgag 3372
Db 1837 GAGCGCCTCGCTTCCAGAAGGAGCCGCCCTTACTCACCTCGCTCCAGCTGGAGAAC 1896
QY 3373 cgacgacttcccgactgactctctcgc 3400
Db 1897 CACAGGACCCCTCCAGGTCTCCCTCCCCC 1924

RESULT 13
US-09-024-429-4
; Sequence 4, Application US/09024429
; Patent No. 6185768
; GENERAL INFORMATION:
; APPLICANT: TSUSAKI, Keiji
; APPLICANT: KUBOTA, Michio
; APPLICANT: SUGIMOTO, Toshiyuki
; TITLE OF INVENTION: RECOMBINANT THERMOSTABLE ENZYME FOR
; TITLE OF INVENTION: CONVERTING MALTOSE INTO TREHALOSE
; NUMBER OF SEQUENCES: 17
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: BROWDY AND NEIMARK
; STREET: 419 Seventh Street, N.W., Suite 300
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20004
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/024,429
; FILING DATE:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/537,002
; FILING DATE: 29-SEP-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 260984/1994
; FILING DATE: 01-OCT-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 255829/1995
; FILING DATE: 08-SEP-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: YUN, Allen C.
; REGISTRATION NUMBER: 37,971
; REFERENCE/DOCKET NUMBER: TSUSAKI-1b
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-628-5197
; TELEFAX: 202-737-3528
; TELEX: 248633
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2889 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
US-09-024-429-4

```

```

Query Match          1.0%;      Score 50.4;  DB 4;  Length 2889;
Best Local Similarity 48.0%;      Pred. No. 0.012;
Matches 215; Conservative 0; Mismatches 221; Indels 12; Gaps 2;

```


ADDRESSEE: BROWDY AND NETWORK
STREET: 419 Seventh Street, N.W., Suite 300
City: Washington
STATE: D.C.
COUNTRY: USA
ZIP: 20004
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/863,010
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/537,002
FILING DATE: 29-SEP-1995
APPLICATION NUMBER: JP 260984/1994
FILING DATE: 01-OCT-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP No. 6087146 yet received
FILING DATE: 08-SEP-1995
ATTORNEY/AGENT INFORMATION:
NAME: BROWDY, Roger L.
REGISTRATION NUMBER: 25,618
REFERENCE/DOCKET NUMBER: TSUSAKI-1
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202-628-5197
TELEFAX: 202-737-3528
TELEX: 248633

```

/ 240353
/ INFORMATION FOR SEQ ID NO: 5:
/   SEQUENCE CHARACTERISTICS:
/     LENGTH: 3600 base pairs
/     TYPE: nucleic acid
/     STRANDEDNESS: DOUBLE
/     TOPOLOGY: linear
/     MOLECULE TYPE: genomic DNA
/     ORIGINAL SOURCE:
/     ORGANISM: Thermus aquaticus
/     INDIVIDUAL ISOLATE: ATCC 33923
/     FEATURE:

```

Query Match 1.0%; Score 50.4; DB 3; Length 3600;
Best Local Similarity 48.0%; Pred. No. 0.013;
Matches 215; Conservative 0; Mismatches 221; Indels 12; Gaps 2;

QY	2956	g c g g c c c g a a g g a g g c c t g c a g c c c a c a a c c a g a a c g g a t c t t c t c o a a c g g c	3015
Db			
Db	2026	G A C C T C C C C T T G G A G G C C T A C C A A G G C C T C G T C C C C T G G A G C T C T T T C G C A G C A A C C C	2085
QY	3016	a g a a g g c g a g a a g c g g g g c c t g g c c g c c a a c c c g g g g a g c c g c a c a g a c c a c t c g t c	3075
Db			
Db	2086	T T C C C C C C G S T G G A G G G G C G T A C C G C T T G A C C T T G G C C C C C A C G G C T T G C C C C T T T C	2145
QY	3076	t c c c g y c t g t g g a c g a g g g c g g c t g g c c g g a g a g t c t g t g c c t c g t g g g a g g c g a g	3135
Db			
Db	2146	G C C C T G A A G C C C G T G G A G G C G T G C T C A C C T C C C C T C C C C G A C T G G G C G A G A G C C C	2205
QY	3136	g a c t c g c g g a a a g a a c a c g a c c t c g a g g a g a c c c g a g t c a g a g c a a g t g c t c c t g	3195
Db			
Db	2206	G C C C C G A G G A G C C G A C T G C C C G G G T C C A C A T G C C C G G G G G C G A G T C T C T C T T G	2265

Qy	3196	caoctgccttcacattggtgaccgagcagcgcgccctggacgacatcctctggggaacctc	3355
Db	2266	g-----TGGACACCCCTGTCTCCACGAAAGGGGGCGGAGGAGCTCTTAACGCCCTC	2316
Qy	3256	tcccagcagcccgagagctgcgcacatctctacagcaagcact- --ggtggccccagctg	3312
Db	2317	GCCACAGACCTTGAAAGGAGAAGAGCTGGTTCGCGCTTCAAGCGCGCAGAGGTGGCCCTCTCTG	2376
Qy	3313	gccagaggatcttcgcgacgcactggagcaccaggacacgcctctgaagccctcggag	3372
Db	2377	GAGCGCCCTCGGCTTCAGAAGGACCGCGCCCTTTACCTCACCCCTGCTCCAGCTGGAGAAC	2436
Qy	3373	cgaggactccccagtgactctctcgc	3400
Db	2437	CACAGAGACCTTCAGGTCTCCCTCCCCC	2464

Search completed: July 24, 2002, 04:24:35
Job time: 18829 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 23, 2002, 22:34:46 ; Search time 3724.92 Seconds
(without alignments)
19164.276 Million cell updates/sec

Title: US-10-044-807-1
Perfect score: 5289
Sequence: 1 atggaatgctgcgtggcgc.....gaacttggtgcaaacgctga 5289

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 13736207 seqs, 6748477542 residues

Total number of hits satisfying chosen parameters: 27472414

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST:*

1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estnu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_hic:*
9: gb_estl:*
10: gb_est2:*
11: gb_hic:*
12: gb_gss:*
13: em_gss_hum:*
14: em_gss_inv:*
15: em_gss_pln:*
16: em_gss_vrt:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	ID	Description
1	824	15.6	1009	10	BM476141
2	623.2	11.8	850	10	BE888902
3	562.2	10.6	568	9	A1917724
4	521	9.9	522	9	A1342006
5	514.4	9.7	581	10	BT523618
6	508	9.6	509	9	A1261611
7	502	9.5	622	9	B8633755
8	478	8.9	979	11	AK020115
9	471.8	8.9	633	9	B8620571
10	456.8	8.6	660	9	B8659292
11	442.6	8.4	700	9	B8643318
12	442.6	8.4	449	10	BF058634
13	414.8	7.8	419	9	A1750970
14	413.8	7.8	619	9	B8478282
15	409.4	7.7	555	9	B8615232
16	407.6	7.7	637	9	B8612331
17	406.2	7.7	586	10	BF510853

c	18	398.8	7.5	454	10	N33407
	19	386.2	7.3	550	10	BE628435
	20	384.4	7.2	643	9	BE193444
c	21	380	7.2	380	9	AI825646
c	22	375.4	7.1	644	10	BF111214
c	23	365.8	6.9	423	9	AA482392
c	24	350.4	6.6	631	10	BF445136
c	25	338	6.4	600	10	BE631939
c	26	327.8	6.2	628	9	AI459225
	27	288	5.4	846	11	AK020361
	28	285	5.4	313	10	W47220
c	29	277.6	5.2	539	9	AW120520
	30	271	5.1	442	9	B8839998
	31	268	5.1	606	10	BE750735
c	32	266	5.0	452	9	AW120794
c	33	264.2	5.0	485	9	AA707140
c	34	263.2	5.0	302	9	AA284278
c	35	247.2	4.7	437	10	W47029
c	36	238.2	4.5	534	9	AA971699
	37	232	4.4	233	10	N44869
c	38	217	4.1	223	10	BE501165
c	39	206	3.9	742	12	AG029671
	40	200.4	3.8	425	10	BF767952
	41	187	3.5	444	12	AQ354354
	42	174.8	3.3	605	9	AI905640
c	43	174.4	3.3	332	9	AV161151
	44	164	3.1	164	10	BF932251
	45	160	3.0	571	9	AI905601

ALIGNMENTS

RESULT 1

BM476141

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

FEATURES

source

BM476141 1009 bp mRNA linear EST 05-FEB-2002
AGENCOURT_647885 NIH_MGC_88 Homo sapiens cDNA clone IMAGE:558669
5', mRNA sequence.
BM476141 GI:18525183
EST.
human.
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 1009)
NIH-MGC http://mgc.nci.nih.gov/
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgabbs@mail.nih.gov
Tissue Procurement: ATCC
cDNA Library Preparation: Life Technologies, Inc.
DNA Sequencing by: The I.M.A.G.E. Consortium (LNL)
Clone distribution: Agencourt Bioscience Corporation
found through the I.M.A.G.E. Consortium/LNL at:
http://image.lnl.gov
Plate: LLAM12282 row: a column: 06
High quality sequence stop: 647.
Location/Qualifiers
1..1009
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:558669"
/clone_lib="NIH_MGC_88"
/tissue_type="duodenal adenocarcinoma, cell line"
/note="Organ: small intestine; Vector: pCMV-SPORT6;
Site 1: NotI; Site 2: SalI; Cloned unidirectionally;
oligo-dr primed. Average insert size 1.767 kb. Library
enriched for full-length clones and constructed by Life
Technologies. Note: this is a NIH_MGC Library."

NI33407 YY41f11.sl
BE628435 uu09f09.Y
BI193444 BB193444
AI825646 wb75b10.x
BF111214 7n44e07.x
AA482392 zv05h07.r
BF445136 nad2la04.
BE631939 uu09f09.x
AI459225 tk11b03.x
AK020361 Mus muscu
W47220 zc39a07.r1
AW120520 UI-M-BH1-
BB839998 BB839998
BG750735 602706924
AW120794 UI-M-BH2.
AA707140 zj33e01.s
AA284278 zc39a07.T
W47029 zc39a07.sl
AA971699 op95a12.s
N44869 YY40f11.r1
BE501165 7a39b04.x
AG029671 Pan trogl
BF767952 CML-CN006
AQ354354 CITBI-El-
AI905640 CM-BT094-
AV161151 AV161151
BF932251 IL2-WT019
AI905601 CM-BT094-

2

361 ctcaagtcgcaagccaaagcaaacctggtgttgtaactagacacctaaggtcttagat 420
 |||||
 Db CTCAAGTCGCAAGCCAAAGGCAACACCTGTTGTTGAACAGGACCTAAGGCTCTAGAT 463
 |||||
 QY ggtacggtgtctatcacagaatcttggatagtgcacacagtggtttatgccaattgtt 480
 |||||
 Db GGTACGGC-TGCTATACAGATCTTGGATATGTCATCAGTGGTTATGCCAAATGTT 522
 |||||
 QY ggtcgcataccagctgggaagcaccgtcaaggaataactgtgggtctgcaacgga 540
 |||||
 Db GGTGCGATACACAGCTGGGAAGCAGCCGCTCAAGGAAGATAACTGTGGGTCTGCAACGGA 582
 |||||
 QY gat-gggfccaacctgcggtgtccagagggcagatataatcccagctctccgcaacaa 599
 |||||
 Db GATGGGTCCACCTTCCGCGCTGGTCCGAGGCGAGTAAATCCAGAGCTCTCAGCAACCCA 642
 |||||
 QY atc-ggatgatactgtgtgtgc-aattccctatggaagttagacatatcgcttgtetta 657
 |||||
 Db ATCGGGATGATCTGTGTTGCAATTCCTATGGGAAGTTGACTATTTCGCTTGGTTTA 702
 |||||
 QY aaaggtcctgatcacttatctctggaacccaaacccctccaggggactaaaggtgaaac 717
 |||||
 Db AAAGGGCTGATCATTCTTTATTGGGAACCAAAA-CTCCACGGGACTCAAGGTGGCAAC 761
 |||||
 QY agtctcagctccacag 733
 |||||
 Db AGGTCATCAACAG 777

RESULT 3
 A1917724/c
 LOCUS
 DEFINITION
 t11c08.x1 NCI_CGAP_GC6 Homo sapiens cDNA clone IMAGE:2240462 3'
 similar to WP:F25H8.3 CE05729 THROMBOSPONDIN LIKE ;, mRNA sequence.
 A1917724
 ACCESSION
 VERSION
 KEYWORDS
 EST.
 A1917724.1 GI:5637579

SOURCE
 human.
 ORGANISM
 Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
 1 (bases 1 to 568)
 NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
 National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
 Tumor Gene Index
 Unpublished (1997)
 Contact: Robert Strausberg, Ph.D.
 Email: cgapbs-re@mail.nih.gov
 Tissue Procurement: Christopher A. Moskaluk, M.D., Ph.D., Michael
 R. Emmert-Buck, M.D., Ph.D.
 cDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima
 Bonaldo, Ph.D.

cDNA Library Arrayed by: Greg Lennon, Ph.D.
 DNA Sequencing by: Washington University Genome Sequencing Center
 Clone distribution: NCI-CGAP clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:
www-bio.llnl.gov/bbrp/image/Image.html
 Insert length: 922 Std Error: 0.00
 Seq primer: -40UP from Gibco
 High quality sequence stop: 437.
 Location/Qualifiers
 1..568

FEATURES
 source
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:2240462"
 /clone_lib="NCI_CGAP_GC6"
 /tissue_type="pooled germ cell tumors"
 /lab_host="DH10B"
 /note="Vector: p773D-Pac (Pharmacia) with a modified
 polylinker; Site_1: Not I; Site_2: Eco RI; plasmid DNA
 from the normalized library NCI_CGAP_GC4 was prepared, and
 ss circles were made in vitro. Following HAP purification,

this DNA was used as tracer in a subtractive hybridization
 reaction. The driver was PCR-amplified cDNAs from a pool
 of 5,000 clones made from the same library (clonoids
 1257096-1258631, 1469064-1470983, and 1475592-1476743).
 Subtraction by Bento Soares and M. Fatima Bonaldo.
 BASE COUNT 120 a 157 c 169 g 121 t 1 others
 ORIGIN

Query Match 10.6%; Score 562.2; DB 9; Length 568;
 Best Local Similarity 99.3%; Pred. No. 3.3e-105;
 Matches 564; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
 QY 1996 tgccacgcaagtgaggaaattggcaagtgagtcacatgtgggtcgccg 2055
 |||||
 Db 568 TGCCACGCAAGGTGGGANATTTGGCAAGTGGAGTCCATCTACATGTGCGGTGCGC 509
 |||||
 QY 2056 ctacagaccagagcgtctctcgcagccacctgtttccagagagatgaatgaacagtc 2115
 |||||
 Db 508 CTACAGACCAGAGAGCTTCTTGACGCCACCTGCTGCCAGAGAGATGAATGAACAGTC 449
 |||||
 QY 2116 atcctggtgatgagctgtgtcgcagcccaagccacgacggtgcaagcttgaacgcg 2175
 |||||
 Db 448 ATCCTGGCTGATGAGCTGTGTGCCACGCCAAGCCAGCAGCGGTGCAAGCTTGTAAACCGC 389
 |||||
 QY 2176 tttaattgccccccagcctggtaccctgcacagtgccagcgtgtccagaacgtggc 2235
 |||||
 Db 388 TTTAATTGCCCCCGAGCTGTGTACCTTGCACAGTGGCAGCGCTTTCAGAACGTGTGGC 329
 |||||
 QY 2236 gggggtgttcagaaacgtgaggtcttcttgaagcagcgcagtgctgatgagcagcttctg 2295
 |||||
 Db 328 GGGGGTGTTCAGAAACGTGAGGTTCTTTGCAAGCAGCGCATGGCTGATGGCAGCTTCTCTG 269
 |||||
 QY 2296 gaggctcctgagacctctctgttcagcttcaaaacctgcctgccagcagcagcagcagcagc 2355
 |||||
 Db 268 GAGCTTCTTGAGACCTTCTCTGTTCAGCTTCAAACTTCCCTGCCAGCAGCATGCAAGAA 209
 |||||
 QY 2356 gatgactgtccagcagagtggtctctctcagactggacagtggtccacaagctgcggg 2415
 |||||
 Db 208 GATGACTGTCCACGAGTGGCTTCTCTCAGACTGGACAGAGTGTCCACAGCTGCGGG 149
 |||||
 QY 2416 gaaggcaccacagactcgaagcgcatttccgaaagatgctgaaacccgctctcaacg 2475
 |||||
 Db 148 GAAGCACCAGACTCGAAGCGCCATTTCGGAAGATGCTGAAACCCGCGCTCTCAACG 89
 |||||
 QY 2476 gttgtcaattccacctgtgccgcgccccctgtctctctctctccatcagccctgtatg 2535
 |||||
 Db 88 GTGTCAATTCCACCTGTGCCGCGCCCTGCTTCTCTCTCTCCATCAGGCCCTGTATG 29
 |||||
 QY 2536 ctggcaacctgtgcaaggcccgccggcgcc 2563
 |||||
 Db 28 CTGGCAACCTGTGCAAGGCACGCGCGCG 1

RESULT 4
 A1342006/c
 LOCUS
 DEFINITION
 qs98d10.x1 NCI_CGAP_GC4 Homo sapiens cDNA clone IMAGE:1946131 3'
 similar to WP:F25H8.3 CE05729 THROMBOSPONDIN LIKE ;, mRNA sequence.
 A1342006
 ACCESSION
 VERSION
 KEYWORDS
 EST.
 A1342006.1 GI:4078933
 SOURCE
 human.
 ORGANISM
 Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
 1 (bases 1 to 522)
 NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
 National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
 Tumor Gene Index
 Unpublished (1997)
 Contact: Robert Strausberg, Ph.D.
 Email: cgapbs-re@mail.nih.gov

Tissue Procurement: Christopher A. Moskaluk, M.D., Ph.D., Michael
Emmert-Buck, M.D., Ph.D.
cDNA Library Preparation: M. Bento Soares, Ph.D.
cDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
www-bio.llnl.gov/bbrp/image/image.html
Seq primer: -400P from Gibco
High quality sequence stop: 482.

FEATURES
source
1. 522
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:1946131"
/clone_lib="NCI-CGAP-CC4"
/tissue_type="pooled germ cell tumors"
/lab_host="DH10B"
/note="Vector: pT7T3D-Pac (Pharmacia) with a modified
polylinker; 1st strand cDNA was prepared from 3 pooled
germ cell tumors, and was then primed with a Not I -
oligo(dT) primer. Double-stranded cDNA was ligated to Eco
RI adaptors (Pharmacia), digested with Not I and cloned
into the Not I and Eco RI sites of the modified pT7T3
vector. Library is normalized. Library was constructed by
Bento Soares and M. Fatima Bonaldo."
BASE COUNT 106 a 139 c 166 g 110 t
ORIGIN

Query Match 9.9%; Score 521; DB 9; Length 522;
Best Local Similarity 99.8%; Pred. No. 9.4e-97;
Matches 521; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2076 ctgagccactgtctccagagagatgaatgaacagtcactcgtgctgagctgtg 2135
Db 522 CTGAGCCACTGCTTCCAGAGATGAATGAACAGTCATCTGCTGATGAGCTGTG 463
QY 2136 tcgcccagccaaagccagcagctgaagctgttaaccgtcttaattgccccagcctg 2195
Db 462 TCGCCAGCNCNAAGCCAGCAGCGTGTGAACCGCTTTAAATGCCCCCAGCCTG 403
QY 2196 gtacctgcacagtggcagcgtgttccagaacgctgtggcggtgttcagaaacgtga 2255
Db 402 GTACCTTGCACATGGCAGCGGTGTTCACAAACGCTGTGGCGGGGTCTTCAGAAACGTGA 343
QY 2256 ggtttcttgcagcagcagcagctgtgctgagcttctgagcttctgagaccttctg 2315
Db 342 GGTCTTTGCAAGCAGCGCATGGCTGTATGGCAGCTTCTTGAGACTCTTCTG 283
QY 2316 ttcagcttcaaacctgctgcccagcagcatgcaagaagatgactgtcccagcagctg 2375
Db 282 TTCAGCTTCAAAACCTGCCGCCACCAAGCATGCAAGAAAGATGACATGTCGCCAGGAGTG 223
QY 2376 gcttctctcagactggacagagtgttccacaagctggtggggaagcaccacagctgaag 2435
Db 222 GCTTCTCAGACTGGACAGAGTGTTCACAAAGCTGCGGGGAAGCACCAGACTCGAAG 163
QY 2436 cgccattgccaagatgctgaaacagcgtctcactcaagctgttcaattccacctgtg 2495
Db 162 CGCCATTTGCCGAAAGATGCTGAAACCGGCTCTCAACGGTGTGTAATTCACCCCTGTG 103
QY 2496 ccgcccctgcttctcttctccatcagcctgtatgctggcaacctgtgcaagcc 2555
Db 102 CCGGCCCTTGCCTTCTCTTCTCTCCATCAGGCCCTGTATGTCGCAACCTGTGCAAGGCC 43
QY 2556 cgggaggccatcccaagagcacagccgcacatcgccgc 2597
Db 42 CGGGGGCCATCCACAGAGCACAGCCGCCACATCGCGCGC 1

RESULT 5
BI523618

LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM

REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

FEATURES
source

BI523618 581 bp mRNA linear EST 29-AUG-2001
603051757F1 NIH_MGC_122 Homo sapiens CDNA clone IMAGE:5201199 5',
mRNA sequence.
BI523618
BI523618.1 GI:15348410
EST.
human.
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
NIH-MGC http://mgs.nci.nih.gov/.
1 (bases 1 to 581)
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue procurement: Life Technologies, Inc.
cDNA Library Preparation by: The I.M.A.G.E. Consortium (LLNL)
cDNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLAM11504 row: b column: 16
High quality sequence stop: 577.
Location/Qualifiers
1. 581
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:5201199"
/clone_lib="NIH_MGC_122"
/lab_host="DH10B"
/note="Organ: pooled lung and spleen; Vector: pCMV-SPORT6;
Site_1: NotI; Site_2: EcoRV (destroyed); RNA source
anonymous pool of 24 week female lung, 16 week female
spleen, and 20-22 week male spleens. Library is oligo-dT
primed and directionally cloned (EcoRV site is destroyed
upon cloning). Average insert size 1.4 kb, insert size
range 1-3 kb. Library is normalized and enriched for
full-length clones and was constructed by C. Gruber
(Invitrogen). Research Genetics tracking code 026. Note:
this is a NIH_MGC Library."
BASE COUNT 140 a 151 c 159 g 131 t
ORIGIN

Query Match 9.7%; Score 514.4; DB 10; Length 581;
Best Local Similarity 99.8%; Pred. No. 2.2e-95;
Matches 515; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 atggaatgctccgtcggggaactcctgtgcacactgctcctcttctggtcttcgtgc 60
Db 66 ATGGAATGCTCCGTCGGGCAACTCCTGGCACACTGCTCTCTTTCTGCTTCTGCTC 125
QY 61 ctgagttccagaccgacgctccgagagagaccggagcgcctatggagtcctggggc 120
Db 126 CTGAGTTCCAGGACCGCAGCTCCGAGGAGGACCGGAGCGCTATGGGATCCCTGGGCG 185
QY 121 ccattgagtgaatgctcacgacacctgcgggggtgggggctctactctctctgaggcgctgc 180
Db 186 CCATGGTGAATGCTCAGCACCTCGGGGGGAGGGGCTCTACTCTCTGAGGGCGTGC 245
QY 181 ctgagcagaagctgtgaagaagaataatccatcacagacatgcagtaattggac 240
Db 246 CTGAGCAGCAAGAGCTGTGAAGGAATAATCCGATACAGAACATGCAGTATGAGGAC 305
QY 241 tccccaccagaagcaggtgatttccgagctcagcaatgctcagctcgaatgctcaag 300
Db 306 TCCCCACCAAGAGCAGGTGATTTCCGAGCTCAGCAATGCTCAGCTCATATGATGCAAG 365
QY 301 caccatggccagttttatgaatggcttctctgtgttaagtgaacctgaacccatttca 360
Db 366 CACCATGGCCAGTTTATGAATGGCTTCTGTGTCTAATGACCTGACACCCATGTTCA 425

	Query Match	9.5%	Score 502;	DB 9;	Length 622;
	Best Local Similarity	87.9%;	Prod. No. 8e-93;		
	Matches 547;	Conservative 0;	Mismatches 75;	Indels 0;	Gaps 0;
Qy	36	gctccttttgcgttcctgcctcctgagttccagagccgcacgctccgagaggacgcg	95		
Db	1	GCTGCTCGTGTGGCTTTCCTACTCTGAGTTCAGAGACTGCACGCTCAGAGGAGACCG	60		
Qy	96	ggacgcgctatgggtagctgcctggggcccatggagtggaatgctcacgcacactcgcgggggtgg	155		
Db	61	TGAAGGCCCTCGGATGCTCGGGGCCCTTGGAGCGAGTGTTCACGAACCTGTGGTGGGG	120		
Qy	156	ggcctctactctctgagcgctgcctgagcagcagaagagctgtgaaggagaagaatccg	215		
Db	121	TGCCCTCCTATTCCCTCAGACCGCTGCCCTGAGCAGCAAGAGCTGTGAAGGGAGAAATATCG	180		
Qy	216	atacagaacatcagtaatgtgagactgccaccacgaagcaggtgattccgcagctcagca	275		
Db	181	ATATAGAACATCGAGTAATGTGGACTGCCACCAGAGAGAGGTGATTTCCGAGCTCAACA	240		
Qy	276	atgctcagctcctaataatggtcaagcaccatgccagtttttatgaatggcttctctgttc	335		
Db	241	GHTGTTCTGCTCATACACGCGTCAAGTACCATGCACAGCTTATGAATGGCTTCTCTGATC	300		
Qy	336	taatgacctgcacaacccatgttcatctcaagtgccaaaggaacacacctggttgt	395		

gagttccaggaccgcacgctccgagaggaccgggacggcctaagggatgcctggggc 120

CTGAGTTCACGAGCTGCACGCTCAGAGGAAGACCGCTGAAGGCCICTGGGATGCGCTGGGGC 204

ccatggagtgaatgtctacgcacactcgcgggggtggggcctctactctctgagggcgtgc 180

||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| 204

CCTTGGAGCGAGTGTTCAGAAACCTGTGGTGGGGTGCTCTCTATTCCCTGAGACGCTGC 204

ctgagcagcaagagctggaaggagaanaatcccgatcacagacaatgcagtaatgtggac 240

||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| 240

CTGAGCAGCAAGAGCTGTGAAGGGAGAAATATCGATATAGAACAATGCGAGTAATGTGGAC 324

tgccaccagaagcaggtgatttccgagctcagcaatgctcagctcataatgatctcaag 300

||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| 300

TGCCCCACCAAGACAGGTGATTTCCGAGCTCAACAGTGTCTGCTCATAACACGCTCAAG 384

caccatggccagtttatgaatggcttctctgtgtctaataggcctgcgaacccatgttca 360

||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| 360

TACCAATGGACAGCTTTATGAATGGCTCTCTGTATCTAATGACCCGCAATCCATGCTCA 444

ctcaagtgcacaagccaaagacaacccctggttgaactagcaacctaaagctctagat 420

||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| 420

CTCAAGTGCACCAACAAAGAACAGCGCTGGTGTGGAACTAGCACCGAAGGTCCTATAT 504

ggtagcgttgctatcacagaacttttgatattgtgcatacagtcagtggttatgcacaattgt 480

||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| 480

GGTACGCGATGCTATACAGAACTACTGGAATGTGCATCAGCGGCTATGCCAATTTGTT 564

ggctcgatcacagctgggaagcacogtcaagagaataactgtggggtctgcacgga 540

||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| 540

GGCTGTGATCACAGCTGGGAGCAGCTGTCAAAAAGGATAACTGTGGAGTGTGCACCGA 624

gatgggtccacctgcc-ggctgggtccgagggcagtatataatcccgactctcgcgaaccaa 599

||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| 599

GATGCTCAACGTGCCGGGTGTGTCGAGGCGAGTATAAATTTTCAGCTCTCTGCAAGTAA 684

||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| 684

at 601

|||||

AT 686

BB20571 633 bp mRNA linear EST 31-AUG

MB20571 RIKEN full-length enriched, 13 days embryo forelimb Mu-

musculus cDNA clone 5930437A14 5', mRNA sequence.

BB20571

BB20571.1 GI:15397167

EST.

house mouse.

Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; I

1 (bases 1 to 633)

Arakawa,T., Carninci,P., Fukuda,S., Furuno,M., Hanagaki,T., Han-

, Hiramoto,K., Hori,F., Ishii,Y., Ito,M., Kawai,J., Konno,H., K

, M., Koya,S., Matsuyama,T., Miyazaki,A., Nomura,K., Ohno,M.,

Okazaki,Y., Okido,T., Saito,R., Sakai,C., Sakai,K., Sano,H., Sa-

D., Shibata,K., Shinagawa,A., Shiraki,T., Sogabe,Y., Suzuki,H., Sa-

Tagami,M., Tagawa,A., Takahashi,F., Takeda,Y., Tanaka,T., Toya-

Muramatsu,M. and Hayashizaki,Y.

RIKEN Mouse ESTs (Arakawa,T., et al. 2001)

Unpublished (2001)

Contact: Yoshihide Hayashizaki

Laboratory for Genome Exploration Research Group, RIKEN Genomic

Sciences Center(GSC), Yokohama Institute

The Institute of Physical and Chemical Research (RIKEN)

1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Ja

Tel: 81-45-503-9222

Fax: 81-45-503-9216

Email: genome-res@gsc.riken.go.jp,

URL:http://genome.gsc.riken.go.jp/

Carninci,P., Shibata,Y., Hayasaka,N., Sugahara,Y., Shibata,K., I

, M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.

Normalization and subtraction of cap-trapper-selected cDNAs to

sequencing pipeline with 384 multicapillary sequencer. Genome Res. 10 (11), 1757-1771 (2000)
 Konno, H., Fukunishi, Y., Shibata, K., Itoh, M., Carninci, P., Sugahara, Y., and Hayashizaki, Y.
 Computer-based methods for the mouse full-length cDNA encyclopedia: real-time sequence clustering for construction of a nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)
 Kondo, S., Shinagawa, A., Saito, T., Kiyosawa, H., Yamanaka, I., Aizawa, K., Fukuda, S., Hara, A., Itoh, M., Kawai, J., Shibata, K. and Hayashizaki, Y.
 Computational Analysis of Full-Length Mouse cDNAs Compared with Human Genome Sequences. Mamm. Genome. 12, 673-677 (2001)
 Please visit our web site (<http://genome.gsc.riken.go.jp>) for further details.
 e mouse tissues.

FEATURES

source

```

1. .660
Location/Qualifiers
/organism="Mus musculus"
/db_xref="taxon:10090"
/clone_lib="D330029102"
/clone_lib="RIKEN full-length enriched, 13 days embryo"
/issue_type="heart"
/dev_stage="13 days embryo"
/lab_host="DH10B"
/note="Site_1: SalI; Site_2: BamHI; cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. 1st strand cDNA was primed with a primer [5' GAGAGAGAGCGCGCAACCTCGAGTGTGTTTCTTTTNN 3'], cDNA was prepared by using trehalose thermo-activated reverse transcriptase and subsequently enriched for full-length by cap-trapper. Second strand cDNA was prepared with the primer adapter of sequence [5' GAGAGAGATCTCGAGTCTAATTAATTAATCCGCCGCCGCC 3']. cDNA was cleaved with BamHI and XhoI. Vector: a modified pBluescript KS(+) after bulk excision from Lambda FLC I."
BASE COUNT 161 a 163 c 191 g 145 t
ORIGIN

Query Match      8.6%; Score 456.8; DB 9; Length 660;
Best Local Similarity 88.8%; Pred. No. 1.6e-83;
Matches 494; Conservative 0; Mismatches 62; Indels 0; Gaps 0;

QY 1 atggaatgctgcgtcgggcaactctgtgcacactgtctctcttctgttctctgtc 60
Db 105 ATGGAATGCTGCGTGGCGACGCTCTGGCACACGCTGCTGCTTCTGCTTCTACTC 164
QY 61 ctgaattccaggaccgcagctccgaggaggacggcgccctatgggagtcctggggc 120
Db 165 CTGAGTTTCAGGACATGACGCTCGAGAGAGACCTGGAAGGCTCTGGGATGCTGGGCG 224
QY 121 ccattgagtgaatgctcagcacctcggggggtggggctctctctctctctctgagcgctgc 180
Db 225 CCTGGAGCGAGTGTTCAGAACCTGTGTGGTGGGGTGGCTCTCTATTCCTGAGACGCTGC 284
QY 181 ctgacgacgaagagctgtgaaggagaataatccgcatcagacaacgacgaatgtgac 240
Db 285 CTGACGACGAGCTGTGAGGAGGAATAATTCGATATAGAACATGACGATGATGAGAC 344
QY 241 tgcaccaccagcagggtgattccgagctcagcgaatgctcagctcataatgatgtcaag 300
Db 345 TGCCACCAGAGCAGGAGTATTCGAGCTCAACAGTGTCTGCTCTATACGACGCTCAAG 404
QY 301 caccatgcccagtttatgaatgcttctctgtgtctaaagaccctgcacaacccatttca 360
Db 405 TACCATGGACAGCTTATGAATGGTCTCTGTATCTAATGACCCGACCAATCCATGCTCA 464
QY 361 ctcaagtgcgaagcgaaggaacacctggtgttgtaactagcacccttaaggtcttagat 420

```

```

|||||CTCAAGTCCCAAGCAAAAGGACCAAGCCTGGTTGTGGAACTAGACCAAGGCTCTTAGAT 524
Db 465
|||||ggtagcgttctctacacgaatctttgtagatgtgcatcagtggtttatgccaaattgtt 480
QY 421
|||||GGTAGCGGATGCTATACAGAACTCACTGGATATGTGCAATGCGGCTATGCCAAATTTGTT 584
Db 525
|||||ggctgcgcatcaccagctgggaagcaccgtcgaagaagtaactgtggggttgcacacgga 540
QY 481
|||||GGCTGTGATCACCAGCTGGGAGGACCTGTCAAGAAGAACTGTGTGAATGTGCAACGGA 644
Db 585
|||||gatgggtccacctgcc 556
QY 541
|||||GATGGCTCAACGCTGCC 660
Db 645

```

RESULT 11

BB643318

LOCUS

DEFINITION

BB643318 RIKEN full-length enriched, 9.5 days embryo parthenogenote

Mus musculus cDNA clone B130031C01 5', mRNA sequence.

ACCESSION

BB643318

VERSION

BB643318.1

KEYWORDS

EST.

SOURCE

house mouse.

ORGANISM

Mus musculus

REFERENCE

1 (bases 1 to 700)

AUTHORS

Arakawa, T., Carninci, P., Fukuda, S., Furuno, M., Hanagaki, T., Hara, A., M., Koya, S., Matsuyama, T., Miyazaki, A., Nomura, K., Ohno, M., Okazaki, Y., Okido, T., Saito, R., Sakai, C., Sakai, K., Sano, H., Sasaki, D., Shibata, K., Shinagawa, A., Shiraki, T., Sogabe, Y., Suzuki, H., Tagami, M., Tagawa, A., Takahashi, F., Takeda, Y., Tanaka, T., Toya, T., Muramatsu, M. and Hayashizaki, Y.

TITLE

RIKEN Mouse ESTs (Arakawa, T., et al. 2001)

JOURNAL

Unpublished (2001)

COMMENT

Contact: Yoshihide Hayashizaki

Laboratory for Genome Exploration Research Group, RIKEN Genomic

Sciences Center (GSC), Yokohama Institute

The Institute of Physical and Chemical Research (RIKEN)

1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan

Tel: 81-45-503-9222

Fax: 81-45-503-9216

Email: genome-res@gs.riken.go.jp,

URL: <http://genome.gsc.riken.go.jp/>

Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K., Itoh

M., Konno, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.

Normalization and subtraction of cap-trapper-selected cDNAs to

prepare full-length cDNA libraries for rapid discovery of new

genes. Genome Res. 10 (10), 1617-1630 (2000)

Wagui, K., Fujiwaka, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E.,

Watahiki, M., Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsuura

S., Kawai, J., Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and

Hayashizaki, Y.

RIKEN integrated sequence analysis (RISA) system--384-format

sequencing pipeline with 384 multicapillary sequencer. Genome Res. 10 (11), 1757-1771 (2000)

Konno, H., Fukunishi, Y., Shibata, K., Itoh, M., Carninci, P., Sugahara

Y. and Hayashizaki, Y.

Computer-based methods for the mouse full-length cDNA

encyclopedia: real-time sequence clustering for construction of a

nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)

Kondo, S., Shinagawa, A., Saito, T., Kiyosawa, H., Yamanaka, I., Aizawa

K., Fukuda, S., Hara, A., Itoh, M., Kawai, J., Shibata, K. and

Hayashizaki, Y.

Computational Analysis of Full-Length Mouse cDNAs Compared with

Human Genome Sequences. Mamm. Genome. 12, 673-677 (2001)

Please visit our web site (<http://genome.gsc.riken.go.jp>) for

further details.

e mouse tissues.

Location/Qualifiers


```

1..700
/organism="Mus musculus"
/db_xref="taxon:10090"
/clone="B130031C01"
/clone.lib="RIKEN full-length enriched, 9.5 days embryo
parthenogenote"
/tissue_type="parthenogenote"
/dev_stage="9.5 days embryo"
/lab_host="DH10B"
/notes="Site 1: SalI; Site 2: BamHI; cDNA library was
prepared and sequenced in Mouse Genome Encyclopedia
Project of Genome Exploration Research Group in Riken
Genomic Sciences Center and Genome Science Laboratory in
RIKEN, Division of Experimental Animal Research in Riken
contributed to prepare mouse tissues. 1st strand cDNA was
primed with a primer [5']
GAGAGAGAGAGATGCCAGAGACTCTTTTTTTTTTTTTTNN 3', cDNA was
prepared by using trehalose thermo-activated reverse
transcriptase and subsequently enriched for full-length by
cap-trapper. cDNA went through one round of subtraction to
Ro1 = 229.0 second strand cDNA was prepared with the
primer adapter of sequence [5']
GAGAGAGATGATTCGATGATTAATTAATTAATCCCCCCCCCC 3'. cDNA
was cleaved with XhoI and BamHI. Vector: a modified
pBluescript KSI(+) after bulk excision from Lambda FLC 1."
BASE COUNT      171 a      174 c      197 g      158 t
ORIGIN

Query Match      8.4%; Score 443; DB 9; Length 700;
Best Local Similarity 86.1%; Pred. No. 1.le-80;
Matches 502; Conservative 0; Mismatches 80; Indels 1; Gaps 1;

QY 1 atgaatgctgcgcgtgggcaactcctggcacactgctctctcttctggcttctcctgc 60
DB 1 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
40 ATGGAATGCTGCCCTCGGGCAGCTCCCTGGCACACCGCTGCTGCTGCTGGCTTCTACT 99
QY 61 ctgagttccaggaccgcactccgagagagaccgacgcctatggatgcctggggc 120
DB 61 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
100 CTGAGTTCAGGACTGCACGCTCAGAGAGAGACCGTGAAGGCCCTCTGGGATGCTGGG 159
QY 121 ccattggagtgaatgctcagcacacctgcgggggtgggctcctactctctgagcgctgc 180
DB 121 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
160 CTTTGGAGCGAGTGTTCAGAACCTGTGGTGGGGTGCTCTCTATTCCTCAGACGGCTGC 219
QY 181 ctgagcagcaagactgtgaaggagaataatccgatacagagaacatgcagtaattggac 240
DB 181 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
220 CTGAGCAGCAAGAGCTGTGAAGGGGAGAAATATTCGATATAGAACATGCAGTAATGTGAC 279
QY 241 tgcaccaccagaagcaggtgattccagctcagacaatctcagctcataatgatgtcaag 300
DB 241 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
280 TGCCACACAGAGCAGGTGATTCGAGCTCAACAGTGTTCGCTCATACGACGCTCAG 339
QY 301 caccattggccagtttataatggcttctctgtgtataatgacctgcacacccattgtca 360
DB 301 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
340 TACCATGCACAGCTTATGAATGGCTTCTGTATCTAATGACCCCGACAATCCATGCTCA 399
QY 361 ctcaagtcccaagccaaaggaaacacctggttgttaactagaacacctaaaggtcttagat 420
DB 361 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
400 CTCAGTGCACCAAGCAAAAGAACACAGCCGGCTGTGGAACTAGCACCCGAAGGCTCTAGAT 459
QY 421 ggtacgcttctctacagaactcttgatattgtcatcagtggtttatgccaaattggt 480
DB 421 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
460 GGTAGCGCATGCTATPACAGATPACACTGGATATGTCATCAGCGGCTATGCCAAATGTT 519
QY 481 ggctcgtgataccacagctgggaagcaccgtcgaaggagaataactgtggggtctgcacgga 540
DB 481 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
520 GCGTGTGATCACCCACTGGGAGCAGCTGTCAAAGAGGATAACTTGGGAGTGGGCAACGGA 579
QY 541 gatgggtccacctgcgcgtggtgcgaggcagtaataatccc 583
DB 541 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
580 AATGGCT-CACGGGCGGCTGGCCAGGCGAGATAAATTTTAC 621

```



```

Db 209 TGGACAGAGTGTCCACAGCTCGGGGAAGGCACCCAGACTCGAAGCGGCATTGGCGGA 150
QY 2449 aagatgctgaacacggcctctcaacggttgctcaatccacctgtgcccgcctcct 2508
Db 149 AAGATGCTGAAACGGGCATCTCAACGGTGTCAATTCACCCCTGTGCCGCCCTCGCT 90
QY 2509 ttctctccctccatcagccctgtatgctggaacctgtgcaaggcccgggcgccatcc 2568
Db 89 TTTCTTCTCCCTCAATCAGGCGCTGTATGCTGGCAACCTGTGCAAGCGCCGGGGGCCATCC 30
QY 2569 acgaagcacagcccgccacatcgcgccgc 2597
Db 29 ACGAAGCACAGCCCGCACATCGGGCGGC 1

RESULT 13
AI750970
LOCUS
DEFINITION
AI750970 419 bp mRNA linear EST 22-JUN-1999
clone NHTBC_cn06h02 Random, mRNA sequence.
ACCESSION
AI750970
VERSION
AI750970.1 GI:5129234
KEYWORDS
EST.
SOURCE
human.
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
1 (bases 1 to 419)
Jia,L.B., Young,M.F., Touchman,J.W., Bouffard,G.G.,
Beckstrom-Sternberg,S.M., Green,E.D., Powell,J.I., Yang,L.M., Robey
P.G., Hotchkiss,R.N. and Francomano,C.A.
SGAP: The Skeletal Genome Anatomy Project
Unpublished (1997)
Contact: Libin Jia
Medical Genetics Branch
National Human Genome Research Institute
10/10C101, 9000 Rockville Pike, Bethesda, MD 20892-1267, USA
Tel: 301-402-4877
Fax: 301-496-7157
Email: libin@helix.nih.gov
DNA Sequencing and analyses by National Institutes of Health
Intramural Sequencing Center (NISC).
Plate: 06 row: h column: 02
Seq primer: M13Rpl reverse primer (ABI).
FEATURES
source
location/Qualifiers
1..419
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="NHTBC_cn06h02"
/clone_lib="Normal Human Trabecular Bone Cells"
/sex="Female"
/tissue_type="Bone"
/cell_type="Trabecular Bone Cells"
/lab_host="SURE"
/note="Organ: Hip; Vector: pBluescript; Site_1: EcoRI"
BASE COUNT 111 a 112 c 114 g 81 t 1 others
ORIGIN

Query Match 7.8%; Score 414.8; DB 9; Length 419;
Best Local Similarity 99.3%; Pred. No. 5.8e-75;
Matches 416; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 3503 gcaagatctcagcggccagcagctctcagcctcgcctgaggtgtgtcaccacctggggcaga 3562
Db 1 GNACAGCTCAGCGGCCCGCAGCTCTCAGCCTCGAGGTGTCTCACCCACCTGGGCGAGA 60
QY 3563 cgggtggccctgagcaggggacactgaggtgtctctcactgtgagggccatcgccacc 3622
Db 61 CGGTGGCCCTGGCAGCGGGACACTGAGTGTCTCTCTGCTGAGGCCATGGGCCACC 120
QY 3623 caaggcctaccatcagctggggccagggaatgagaagaatttcagttcagtagcaggattc 3682
|||||

```

```

Db 121 CAAGGCTTACCATTGAGCTGGGCCAGGAATGAGGAAGATTTCAGTTAGTCACAGGATTC 180
QY 3683 ttctacagcagatgattccttacagatcttggcaccagtggaagcagatgtgggtttct 3742
Db 181 TTTACAGCCAGATGATTCCTTACAGATCTTGGCACCAGTGGAGCAGATGTGGTTTCT 240
QY 3743 acatttcaatggcaccacatgcttgggatacagactctgtctccattgctccacattag 3802
Db 241 ACATTGCAATGGCACCACCAATGCCCTGGGATACGACTCTGTCTCCATTGCCCTCATATG 300
QY 3803 caggaaagccactagtgaacacgtcacgaatgacagtgcagtgatcaacacgagaaagcctgac 3862
Db 301 CAGGAAAGCCACTAGTCAAAACGTCACGAATGACAGTATCAACACCGAAGGCTGCAG 360
QY 3863 tcacagtcgatattagggaagcaccalcacacagtcgagggagtggaatgtgacaatcaac 3921
Db 361 TCACAGTCGATATAGGAAGCACCATCAAAACAGTGCAGGGAGTGAATGTGACATCAAC 419

```

RESULT 14

```

BB478282 619 bp mRNA linear EST 25-OCT-2001
BB478282 RIKEN full-length enriched, 13 days embryo heart Mus
musculus cDNA clone D330029102 3', mRNA sequence.
BB478282
VERSION
BB478282.2 GI:16440420
KEYWORDS
EST.
SOURCE
house mouse.
ORGANISM
Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE
1 (bases 1 to 619)
AUTHORS
Arakawa,T., Carninci,P., Fukuda,S., Furuno,M., Hanagaki,T., Hara,A.,
Hiramoto,K., Hori,F., Ishii,Y., Ito,M., Kawai,J., Konno,H., Kouda
M., Koya,S., Matsuyama,T., Miyazaki,A., Nomura,K., Ohno,M., Kouda
Okazaki,Y., Okido,T., Saito,R., Sakai,C., Sakai,K., Sano,H., Sasaki
D., Shibata,K., Shinagawa,A., Shiraki,T., Sogabe,Y., Suzuki,H.,
Tagami,M., Tagawa,A., Takahashi,F., Takeda,Y., Tanaka,T., Toya,T.,
Muramatsu,M. and Hayashizaki,Y.
RIKEN Mouse ESTs (Arakawa,T., et al. 2001)
Unpublished (2001)
On Jul 23, 2000 this sequence version replaced gi:9395891.
Contact: Yoshihide Hayashizaki
Laboratory for Genome Exploration Research Group, RIKEN Genomic
Sciences Center(GSC), Yokohama Institute
The Institute of Physical and Chemical Research (RIKEN)
1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
Tel: 81-45-503-9222
Fax: 81-45-503-9216
Email: genome-res@gsc.riken.go.jp,
URL:http://genome.gsc.riken.go.jp/
Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K., Itoh
M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
Normalization and subtraction of cap-trapper-selected cDNAs to
prepare full-length cDNA libraries for rapid discovery of new
genes. Genome Res. 10 (10), 1617-1630 (2000)
wagi,K., Fujiwaka,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E.,
Watahiki,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsumura
S., Kawai,J., Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and
Hayashizaki,Y.
RIKEN integrated sequence analysis (RISA) system--384-format
sequencing pipeline with 384 multicapillary sequencer. Genome Res.
10 (11), 1757-1771 (2000)
Konno,H., Fukunishi,Y., Shibata,K., Itoh,M., Carninci,P., Sugahara
Y. and Hayashizaki,Y.
Computer-based methods for the mouse full-length cDNA
encyclopedia: real-time sequence clustering for construction of a
nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)
Kondo,S., Shinagawa,A., Saito,T., Kiyosawa,H., Yamanaka,I., Aizawa
K., Fukuda,S., Hara,A., Itoh,M., Kawai,J., Shibata,K. and
Hayashizaki,Y.
Computational Analysis of Full-Length Mouse cDNAs Compared with
Human Genome Sequences Mamm. Genome. 12, 673-677 (2001)

```

TITLE
JOURNAL
COMMENT

Please visit our web site (<http://genome.gsc.riken.go.jp/>) for further details.

cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN, Division of Experimental Animal Research in Riken contributed to prepare mouse tissues.

Location/Qualifiers

1. 619

/organism="Mus musculus"

/db_xref="taxon:10090"

/clone="D330029102"

/clone_lib="RIKEN full-length enriched, 13 days embryo heart"

/tissue_type="heart"

/dev_stage="13 days embryo"

/lab_host="DH10B"

/note="Site_1: Sali; Site_2: BamHI; cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN, Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. 1st strand cDNA was primed with a primer [5' GAGAGAGAGCGCGCGCACTCGAGTTTTTTTTTTTTTTTNN 3'], cDNA was prepared by using trehalose thermo-activated reverse transcriptase and subsequently enriched for full-length by cap-trapper. Second strand cDNA was prepared with the primer adapter of sequence [5' GAGAGAGATTCGAGTTAATTAATATCCCGCCCCCCCC 3']. cDNA was cleaved with BamHI and XhoI. Vector: a modified pBluescript KS(+) after bulk excision from Lambda FLC I."

160 a 142 c 169 g 148 t

BASE COUNT

ORIGIN

Query Match

Best Local Similarity 7.8%; Score 413.8; DB 9; Length 619;

Mismatches 451; Conservative 0; Mismatches 62; Indels 0; Gaps 0;

QY 89 aggacgggaagcctatggatgctggggccatggagtgatgctacgacacgtcg 148

DB 1 AAGACCGTGAAGGCTTCTGGGATGCTTGGGGCCCTTGGAGCGAGTGTTCAGAACCTGTG 60

QY 149 ggggtggggcctactctctgagcgtgctgctgagcagcaagagctgtgaaggagaa 208

DB 61 GTGGGGTGCCTCTATTCCTGAGACGCTGCTTGAGCAGCAAGAGCTGTGAAGGGAGAA 120

QY 209 atatccgatacagaacatgcagtaatgtgagctgcccaccagagcaggtgattccgag 268

DB 121 ATATTGATATAGACATGCAGTAATGTGGACTGCCACCAGAACGAGTGTTCGCGAG 180

QY 269 ctgagaatgctcagctcataatgatgtcaagcacatggccagttttatgaatggcttc 328

DB 181 CTCACAGTGTCTGCTCATACACGCTCAAGTACCATGCACAGCTTTATGATGGTTC 240

QY 329 ctgtgtctaataaccctgacaacccatgttcaactcaagtccaagccaaagggaacacc 388

DB 241 CTGTATCTAATACCCGACAAATCCATGCTCACTCAAGTCCCAAGCAAAAGGAACCGCC 300

QY 389 tgggtgtgaactagacacctaaggtcttagatggtgaacgcttgctatacagaatcttgg 448

DB 301 TGGTGTGGAACATAGCACCAGGCTCTTAGATGTGTAGCGATGTATACAGAAATCACTGG 360

QY 449 atattgtcatcagtggtttatgtcaaatgtgtgctgcgacacacagctgggaagcaccg 508

DB 361 ATATGTGCATCAGCGGCTATGCCAAATTTGCTGTGTATCACCAGCTGGGGGACACTG 420

QY 509 tcaaggagaataactgtgggtgtgcaacggagatgggtccacctgcggtggtccgag 568

DB 421 TCAAGAGGATAACTGTGGAGTGTGCAACGGAGATGGCTCAACGTGCGCGGTGTCGCGAG 480

QY 569 ggcagataataatccacgtctccgcacccaat 601

Db 481 GCCAGTATAAATCTCAGCTCTCTGCAAGTAAT 513

RESULT 15

BB615232

LOCUS BB615232

DEFINITION BB615232 RIKEN full-length enriched, adult male testis Mus musculus cDNA clone 4930443P21 5', mRNA sequence.

ACCESSION BB615232

VERSION BB615232.1 GI:16455520

KEYWORDS EST.

SOURCE house mouse.

ORGANISM Mus musculus

REFERENCE 1 (bases 1 to 555)

AUTHORS Arakawa,T., Carninci,P., Fukuda,S., Furuno,M., Hanagaki,T., Hara,A., Hiramoto,K., Hori,F., Ishii,Y., Ito,M., Kawai,J., Konno,H., Kouda,M., Koya,S., Matsuyama,T., Miyazaki,A., Nomura,K., Ohno,M., Okazaki,Y., Okido,T., Saito,R., Sakai,C., Sakai,K., Sano,H., Sasaki,D., Shibata,K., Shinagawa,A., Shiraki,T., Sogabe,Y., Suzuki,H., Tagami,M., Tagawa,A., Takahashi,F., Takeda,Y., Tanaka,T., Toyota,T., Muramatsu,M. and Hayashizaki,Y.

TITLE RIKEN Mouse ESTs (Arakawa,T., et al. 2001)

JOURNAL Unpublished (2001)

COMMENT Contact: Yoshihide Hayashizaki
Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center(GSC), Yokohama Institute
The Institute of Physical and Chemical Research (RIKEN)
1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
Tel: 81-45-503-9222
Fax: 81-45-503-9216
Email: genome-res@gsc.riken.go.jp,
URL:<http://genome.gsc.riken.go.jp/>
Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K., Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes. Genome Res. 10 (10), 1617-1630 (2000)
wagi,K., Fujiwaki,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E., Watahiki,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsuura,S., Kawai,J., Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and Hayashizaki,Y.
RIKEN integrated sequence analysis (RISA) system-384-format sequencing pipeline with 384 multicapillary sequencer. Genome Res. 10 (11), 1757-1771 (2000)
Konno,H., Fukunishi,Y., Shibata,K., Itoh,M., Carninci,P., Sugahara,Y. and Hayashizaki,Y.
Computer-based methods for the mouse full-length cDNA encyclopedia: real-time sequence clustering for construction of a nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)
Kondo,S., Shinagawa,A., Saito,T., Kiyosawa,H., Imanaka,I., Aizawa,K., Fukuda,S., Hara,A., Itoh,M., Kawai,J., Shibata,K. and Hayashizaki,Y.
Computational Analysis of Full-length Mouse cDNAs Compared with Human Genome Sequences. Mamm. Genome. 12, 673-677 (2001)
Please visit our web site (<http://genome.gsc.riken.go.jp/>) for further details.
e mouse tissues.

FEATURES

Location/Qualifiers

1. 555

/organism="Mus musculus"

/strain="C57BL/6J"

/db_xref="taxon:10090"

/clone="4930443P21"

/clone_lib="RIKEN full-length enriched, adult male testis"

/sex="male"

/tissue_type="testis"

/dev_stage="adult"

/lab_host="SOLR"

/note="Site_1: XhoI; Site_2: BamHI; cDNA library was prepared and sequenced in Mouse Genome Encyclopedia

Project of Genome Exploration Research Group in Riken
Genomic Sciences Center and Genome Science Laboratory in
RIKEN, Division of Experimental Animal Research in Riken
contributed to prepare mouse tissues. 1st strand cDNA was
primed with a primer [5'
GAGAGAGAGAGATCCAGAGATCTTTTCTTTTCTTTTNN 3'], cDNA was
prepared by using trehalose thermo-activated reverse
transcriptase and subsequently enriched for full-length by
cap-trapper. Second strand cDNA was prepared with the
primer adapter of sequence [5'
GAGAGAGAGAGAGCGCGCAATTAATTCGAGTAATTAATTAATCCGCCCCCC
3']. cDNA was cloned into the XhoI and BamHI sites. *

BASE COUNT 141 a 150 c 143 g 120 t 1 others
ORIGIN

Query Match 7.7%; Score 409.4; DB 9; Length 555;
Best Local Similarity 88.6%; Pred. No. 8.2e-74;
Matches 443; Conservative 0; Mismatches 57; Indels 0; Gaps 0;

QY 3513 agcgccagcagcagctctcagcctcgaggtggtcaccacctgggacagcagtggtgcccct 3572
Db 1 AGCGGCACAGCAACTCTCAGCCTCGGAGGTAGTCACCCACTGGGCAGACTGTGGCTCT 60

QY 3573 ggcagcggtgacactgagtggtctctgcaactgtgagccatcgccaccaccagggcctac 3632
Db 61 AGCCAGTGGGACCTGAGTGTGTTCTGCACTGTGAGGGGTAGGCAACCCAAAGGCCAC 120

QY 3633 catcagctggccaggaatggagaaagtctcagttcagtgacagagattctctacagcc 3692
Db 121 CATCCACTGGACCGAGCAACGAGACAGTCAATTCAGTGACAGGATTCCTTACAGCC 180

QY 3693 agaatattcttacagattttggcaccagtggaagcagatgtggtttctacacttgcaa 3752
Db 181 AGATGATTCCTTACAGATCTTGGCACCTGGGAAGCTGATGTGGTTCCTATCTGCAA 240

QY 3753 tgccaccatgcttgggtacagactctgtctccattgcccgtcacattagcaggaagcc 3812
Db 241 TGCCACAATGCTTGGGATATGACTCTCTCTCCATTGCTGCACATTAGCAGGAAGCC 300

QY 3813 actagtgaacacgtcacgaatgacagtgatcaacacgagagagcctgcagtcacagtcga 3872
Db 301 ACTAGTGAAGACATCCGAATGACTGTGCTCAACACAGAGAGCCCTACAGTCACAGTGA 360

QY 3873 tataggaagcaccatcaaaacagtcagggagtggaatgtgacaatcaactgccaggttgc 3932
Db 361 TATCGTGGCACCGTCAGAACCGTCCGAGGAGTGAATGTGACAATTAATGCAAGAGGC 420

QY 3933 aggagtgctgaagtcagtcacttgggttcaggaataaaagcaaaactgggtcccccga 3992
Db 421 AGGTGTACTGAAGCTGAAGTCACCTGGTTCAGAAATAAAGCAAAATTGGGCTCCTNCCA 480

QY 3993 ccattgcacgaaggtcct 4012
Db 481 CCACCTGCATGAGGCTCCT 500

Search completed: July 24, 2002, 00:27:47
Job time: 6781 sec

